

Durham Research Online

Deposited in DRO:

18 July 2017

Version of attached file:

Published Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Jones, D. O. and Scolnic, D. M. and Riess, A. G. and Kessler, R. and Rest, A. and Kirshner, R. P. and Berger, E. and Ortega, C. A. and Foley, R. J. and Chornock, R. and Challis, P. J. and Burgett, W. S. and Chambers, K. C. and Draper, P. W. and Flewelling, H. and Huber, M. E. and Kaiser, N. and Kudritzki, R.-P. and Metcalfe, N. and Wainscoat, R. J. and Waters, C. (2017) 'Measuring the properties of dark energy with photometrically classified Pan-STARRS supernovae. I. Systematic uncertainty from core-collapse supernova contamination.', *Astrophysical journal*, 843 (1). p. 6.

Further information on publisher's website:

<https://doi.org/10.3847/1538-4357/aa767b>

Publisher's copyright statement:

© 2017. The American Astronomical Society. All rights reserved.

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.



Measuring the Properties of Dark Energy with Photometrically Classified Pan-STARRS Supernovae. I. Systematic Uncertainty from Core-collapse Supernova Contamination

D. O. Jones¹, D. M. Scolnic^{2,10}, A. G. Riess^{1,3}, R. Kessler², A. Rest³, R. P. Kirshner^{4,5}, E. Berger⁴, C. A. Ortega¹, R. J. Foley⁶, R. Chornock⁷, P. J. Challis⁴, W. S. Burgett⁸, K. C. Chambers⁸, P. W. Draper⁹, H. Flewelling⁸, M. E. Huber⁸, N. Kaiser⁸, R.-P. Kudritzki⁸, N. Metcalfe⁹, R. J. Wainscoat⁸, and C. Waters⁸

¹ Department of Physics and Astronomy, The Johns Hopkins University, Baltimore, MD 21218, USA

² University of Chicago, Kavli Institute for Cosmological Physics, Chicago, IL 60637, USA

³ Space Telescope Science Institute, Baltimore, MD 21218, USA

⁴ Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, MA 02138, USA

⁵ Department of Physics, Harvard University, Cambridge, MA 02138, USA

⁶ Department of Astronomy and Astrophysics, University of California, Santa Cruz, CA 92064, USA

⁷ Astrophysical Institute, Department of Physics and Astronomy, 251B Clipping Lab, Ohio University, Athens, OH 45701, USA

⁸ Institute for Astronomy, University of Hawaii at Manoa, Honolulu, HI 96822, USA

⁹ Department of Physics, Durham University, South Road, Durham DH1 3LE, UK

Received 2016 November 21; revised 2017 May 20; accepted 2017 May 30; published 2017 June 26

Abstract

The Pan-STARRS (PS1) Medium Deep Survey discovered over 5000 likely supernovae (SNe) but obtained spectral classifications for just 10% of its SN candidates. We measured spectroscopic host galaxy redshifts for 3147 of these likely SNe and estimate that ~ 1000 are Type Ia SNe (SNe Ia) with light-curve quality sufficient for a cosmological analysis. We use these data with simulations to determine the impact of core-collapse SN (CC SN) contamination on measurements of the dark energy equation of state parameter, w . Using the method of Bayesian Estimation Applied to Multiple Species (BEAMS), distances to SNe Ia and the contaminating CC SN distribution are simultaneously determined. We test light-curve-based SN classification priors for BEAMS as well as a new classification method that relies upon host galaxy spectra and the association of SN type with host type. By testing several SN classification methods and CC SN parameterizations on large SN simulations, we estimate that CC SN contamination gives a systematic error on w (σ_w^{CC}) of 0.014, 29% of the statistical uncertainty. Our best method gives $\sigma_w^{CC} = 0.004$, just 8% of the statistical uncertainty, but could be affected by incomplete knowledge of the CC SN distribution. This method determines the SALT2 color and shape coefficients, α and β , with $\sim 3\%$ bias. However, we find that some variants require α and β to be fixed to known values for BEAMS to yield accurate measurements of w . Finally, the inferred abundance of bright CC SNe in our sample is greater than expected based on measured CC SN rates and luminosity functions.

Key words: cosmology; observations – dark energy – supernovae: general

1. Introduction

Since the discovery of cosmic acceleration (Riess et al. 1998; Perlmutter et al. 1999), measuring the properties of dark energy with Type Ia supernovae (SNe Ia) has been predicated on the spectroscopic confirmation of SN Ia candidates. However, as the size of individual SN Ia samples surpasses 1000 SNe, obtaining spectra for each Type Ia candidate is becoming prohibitively expensive. Only a small fraction of SNe Ia from current and future surveys such as the Dark Energy Survey (DES) and the Large Synoptic Survey Telescope (LSST) will have spectroscopic classification. Without spectroscopic classification, core-collapse SN (CC SN) contamination can bias our estimates of cosmological parameters (Kunz et al. 2007; Falck et al. 2010).

Without SN spectroscopy, the shape and color of a photometric SN light curve can be used as a less precise diagnostic of the type. Campbell et al. (2013) used SDSS *ugriz* light curves to classify 752 SNe as likely Type Ia, enough to measure the dark energy equation of state parameter, w , with $\sim 10\%$ statistical uncertainty. Their sample was selected from light-curve properties and a classifier that compares each observed light curve to SN Ia and CC SN templates (PSNID;

Sako et al. 2011). Their final sample comprised just 3.9% CC SNe. While Campbell et al. (2013) has the *only* SN Ia-based measurement of w to date that does not use spectroscopic classification for its SNe, the measurement did not include systematic uncertainties. In addition, contaminating CC SNe bias their measurements of SN Ia dispersion and the correlation between SN luminosity and light-curve rise/decline rate by $\sim 60\%$.

Many light-curve classifiers use the “naïve Bayes” approximation, which assumes that all observables that indicate SN type are uncorrelated. Machine learning techniques can often outperform these classifiers, yielding higher SN Ia classification efficiency (the fraction of SNe Ia classified correctly) and lower CC SN contamination (Lochner et al. 2016; Möller et al. 2016). On SDSS-SN data, the Sako et al. (2014) kd-tree nearest neighbor (NN) method has a purity comparable to that of Campbell et al. (2013) but accurately classifies ~ 1.4 times as many real SNe Ia in a given sample.

An important caveat is that nearly all classifiers are optimized on simulations with little evaluation on real data. Simulations, in turn, depend on CC SN templates and knowledge of the CC SN luminosity functions (LFs) and rates. CC SNe are diverse, far more so than SNe Ia, and only a limited number of high-quality templates are publicly available. Training a classifier directly on survey data is possible but

¹⁰ Hubble, KICP Fellow.

can be sub-optimal due to limited numbers of CC SNe observed and the dependence of classifier results on the specific survey characteristics (e.g., observing cadences, filters, and signal-to-noise ratios (S/N)).

We can make SN classification less dependent on CC SN templates, LFs, and rates by incorporating host galaxy data. Because many SNe Ia have a $\gtrsim 1$ Gyr delay time between progenitor formation and explosion (Rodney et al. 2014), they are the only type of SNe found in early-type galaxies (with very few known exceptions; Suh et al. 2011). Foley & Mandel (2013) found that it was possible to accurately classify the $\sim 20\%$ of SNe Ia found in elliptical galaxies if the morphology of their host galaxy is known.

Though these results are encouraging, light-curve and host galaxy classification alone may not be enough to enable a measurement of w as precise as measurements using spectroscopically classified SNe (e.g., Betoule et al. 2014, $w = -1.027 \pm 0.055$). A difference in w of 5% corresponds to a change of 0.02 mag from $z = 0$ to $z = 0.5$; if CC SNe are 1 mag fainter than SNe Ia on average, a bias of 0.02 mag can be induced by just a 2% CC SN contamination in a high- z sample such as PS1. If the contaminating distribution of CC SNe is more than 1 mag fainter (this depends on survey Malmquist bias), it takes even fewer CC SNe to bias w by an equivalent amount.

A Bayesian method, however, could use the probabilities that SNe are of Type Ia as priors to simultaneously determine distances to Ia and CC SNe without bias. We refer to this method as Bayesian Estimation Applied to Multiple Species (BEAMS) following Kunz et al. (2007, hereafter KBH07; see also Press 1997 and Rubin et al. 2015). KBH07 test BEAMS on a simplistic SN simulation and find that it gives near-optimal accuracy and uncertainties on SN Ia distances.

Hlozek et al. (2012) test BEAMS further with Monte Carlo simulations of the Sloan Digital Sky Survey SN survey (SDSS-SN; Frieman et al. 2008; Kessler et al. 2009a). BEAMS biases measurements of the cosmic matter and dark energy densities, Ω_M and Ω_Λ , by less than the statistical uncertainties measured from their simulations. Their results demonstrated that SDSS SNe without spectroscopic classification can significantly improve cosmological constraints relative to the SDSS spectroscopic sample (Kessler et al. 2009a). Hlozek et al. (2012) did not measure the systematic uncertainties from their method.

As with SDSS, Pan-STARRS (PS1) discovered far more SNe Ia than could be observed spectroscopically. Spectroscopically confirmed SNe Ia from the first $\sim 1/3$ of PS1 have been used to measure cosmological parameters but constitute only a small fraction of the available data (Rest et al. 2014, hereafter R14; Scolnic et al. 2014b). In this study, we use PS1 SNe with and without spectroscopic classification as a tool for testing SN classifiers, understanding CC SN contaminants, and measuring the systematic error due to CC SN contamination. In total, PS1 has 1145 SNe with high-quality light curves and spectroscopic redshifts—both host galaxy and SN redshifts—that can be used to measure cosmological parameters (including a \sim few percent CC SN contamination). Here, we focus on the 1020 likely SNe Ia with spectroscopic host galaxy redshifts, 143 of which are spectroscopically confirmed, in order to study a sample with fewer selection biases (Section 2.1).

The goal of this study is to develop the methods necessary to measure cosmological parameters robustly using PS1 SNe without spectroscopic classifications (hereafter referred to as

photometric SNe). Our full cosmological results from these data will be presented in a future analysis.

In Section 2, we present the sample and our host galaxy redshift follow-up survey. Section 3 discusses our SNANA simulations of the PS1 sample and our assumptions about the CC SN population. Section 4 describes our Bayesian parameter estimation methodology. In Section 5, we test BEAMS on simulations and subsamples of PS1 photometric SNe. In Section 6, we test the robustness of these results by exploring several variants of the method. The uncertainties in our simulations and methodology are discussed in Section 7 and our conclusions are in Section 8.

2. The Pan-STARRS Photometric Supernova Sample

The Pan-STARRS medium deep survey covers ten 7 square degree fields in five broadband filters, with typical $griz_{P1}$ observational cadences of 6 images per 10 days and a 5 day gap during bright time during which y_{P1} images are taken. Typical 5σ detection limits are ~ 23 AB mag for $griz_{P1}$, albeit with significant variation. For a complete description of the PS1 survey, see Kaiser et al. (2010) and R14.

PS1 images are processed using an image subtraction pipeline that is described in detail in Rest et al. (2005) and R14. To measure final light curves for the PS1 photometric sample (and the full spectroscopic sample; D. M. Scolnic et al. 2017, in preparation), we made several improvements to that pipeline. We more than doubled the typical number of images that are combined to create a deep template for subtraction, we refined our method of selecting stars to build the point-spread function (PSF) model, and we improved the zeropoint calibration. These improvements will be described in detail in D. M. Scolnic et al. (2017, in preparation).

Pan-STARRS discovered 5235 likely SNe during its four years of operation and obtained spectra for 520 SNe. We collected 3147 spectroscopic host galaxy redshifts of these likely SNe (Section 2.1). In addition to SN candidates, we observed spectra for thousands of variable stars, active galactic nuclei (AGN), flaring M dwarfs, and other transients that will be published in future work.

2.1. Host Galaxy Redshift Survey

During the PS1 survey, many SN host redshifts were measured using the Hectospec multifiber instrument on the MMT (Fabricant et al. 2005; Mink et al. 2007). Near the end of PS1 operations, we began an additional survey with Hectospec to obtain redshifts for as many host galaxies as possible. Redshifts were also obtained with the Apache Point Observatory 3.5 m telescope¹¹ (APO), the WIYN telescope,¹² and for the southernmost PS1 field, the Anglo-Australian Telescope (AAT). We chose candidate host galaxies for follow-up in a largely unbiased way; we did not prioritize SNe based on their magnitudes, colors, or whether or not an SN spectrum had previously been obtained. Approximately 600 of our redshifts come from SDSS (Smee et al. 2013) or other public redshift surveys.¹³

¹¹ <http://www.apo.nmsu.edu/arc35m/>

¹² The WIYN Observatory is a joint facility of the University of Wisconsin-Madison, Indiana University, the National Optical Astronomy Observatory, and the University of Missouri.

¹³ We include redshifts from 2dFGRS (Colless et al. 2003), 6dFGS (Jones et al. 2009), DEEP2 (Newman et al. 2013), VIPERS (Scodreggio et al. 2016), VVDS (Le Fèvre et al. 2005), WiggleZ (Blake et al. 2008), and zCOSMOS (Lilly et al. 2007).

We used the galaxy size- and orientation-weighted R parameter to identify the most likely host galaxy for each SN (Sullivan et al. 2006). The isophotal limit of a galaxy corresponds to $R \sim 3$. We use the redshift of the host galaxy with the lowest R if it has $R \leq 5$ following Sullivan et al. (2006). See Gupta et al. (2016) for a similar but more rigorous method of identifying SN host galaxies.

To estimate the fraction of SNe for which we incorrectly determined which galaxy was the host, we compared redshifts derived from the spectroscopic redshifts of SNe to the spectroscopic redshifts of their most likely host galaxies. We found that only 2 of 169 hosts with reliable redshifts had evidence of a host galaxy mismatch, $|z_{\text{SN}} - z_{\text{host}}| > 0.02$. Both of these hosts had multiple large, nearby galaxies with $R < 5$. This mismatch fraction suggests that $1.2\% \pm 0.5\%$ of our redshifts are incorrect due to mismatched hosts.

Compared to spectroscopically confirmed SNe, it is unlikely that photometric SNe have a higher fraction of mismatched hosts. The spectroscopic targeting preferentially followed SNe with a larger separation from the center of their host galaxies or SNe with fainter hosts, as these SNe have spectra with less galaxy light contamination. Just 11% of photometrically classified SNe are outside the isophotal radii of their host galaxies compared to 24% of the 169 SN host pairs. However, we also note that the 169 SN host pairs have preferentially brighter hosts than the full sample and have a median redshift of 0.21 compared to the median redshift of 0.3 for the full sample. It may be somewhat easier to mismatch a host galaxy at high- z as galaxies are more difficult to detect, but we expect this to be a subdominant effect as Gupta et al. (2016) finds the fraction of mismatched hosts to be approximately constant at $z < 0.6$ in a DES-like survey (which has similar depth to PS1 templates).

The other source of incorrect redshifts is the measurement of velocities from host galaxy spectra. We measured redshifts by cross-correlating our spectra with galaxy templates (the RVSAO package; Kurtz & Mink 1998) and visually inspecting the results. Over the course of the survey, we observed over 1500 transient hosts multiple times. For ~ 250 of these hosts, at least one observation yielded a redshift with a high Tonry & Davis (1979, hereafter TDR) cross-correlation parameter ($\gtrsim 9$ –10).

By restricting our sample to hosts with $\text{TDR} > 4$ and redshifts of $0.01 < z < 0.75$, we measure a false redshift fraction of $1.4\% \pm 1.3\%$. At $z > 0.75$, few SNe could be discovered by PS1 or have their host redshifts measured with our program (Figure 1). Including mismatched hosts, the total percentage of incorrect redshifts we expect is $2.6\% \pm 1.4\%$. In Section 3, we simulate this fraction of false redshifts so that this effect will be incorporated in our BEAMS systematic error budget.

In total, we observed 3930 host galaxies and have 3147 reliable redshifts. The telescopes and instruments comprising our redshift survey are summarized in Table 1. Figure 1 shows the r magnitudes, redshifts, and best-fit SED model for the PS1 photometric sample. Eighty-seven percent of PS1 SNe with detectable host galaxies were observed with our redshift follow-up program and reliable redshifts were measured for 73% of those galaxies. We measured redshifts for a large number of both emission-line and absorption-line galaxies. These data have a median redshift of 0.30.

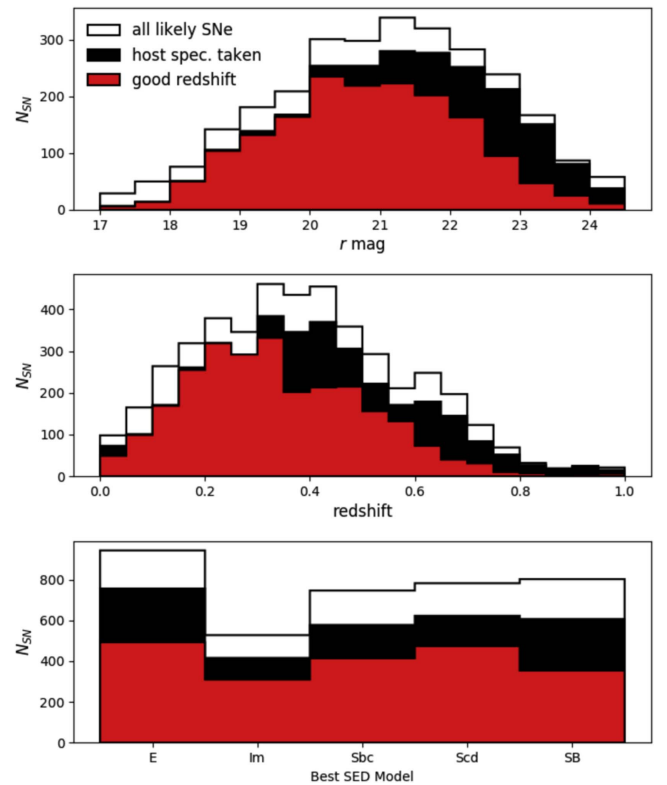


Figure 1. Host properties from PS1 as a function of r mag, redshift, and best-fit SED model. Out of the full sample of 5235 PS1 SNe (white; host galaxy photo- z), we observed 3930 hosts (black; photo- z) and measured accurate redshifts for 3147 (red; spec- z). Our redshift survey has nearly 100% success to $r = 21$ and has a median redshift of 0.30. We obtained redshifts for a large number of both emission-line and absorption-line galaxies.

2.2. SALT2 Selection Requirements

Throughout this work, we use the SALT2.4 model (Guy et al. 2010, implemented in B14) to measure SN light-curve parameters. We use these light-curve parameters to standardize SNe Ia and select the SNe Ia that can best measure cosmological parameters. The Tripp estimator uses SALT2 light-curve parameters to infer the SN distance modulus, μ (Tripp 1998):

$$\mu = m_B + \alpha \times X_1 - \beta \times C - M. \quad (1)$$

m_B is the log of the light-curve amplitude, X_1 is the light-curve stretch parameter, and C is the light-curve color parameter. These parameters are all measured by the SALT2 fitting program, but deriving the distance modulus from them depends on the nuisance parameters α , β , and M . M is degenerate with the Hubble constant, H_0 , and will be marginalized over during cosmological parameter estimation.

To avoid unexpected biases in our sample selection, we use light-curve selection requirements (cuts) from previous analyses using spectroscopically confirmed SNe. We make the same series of cuts to PS1 SN light curves as Betoule et al. (2014) and add one additional cut on the SALT2 fit probability following R14. These cuts include uncertainty-based cuts that ensure the shape and time of maximum light of each SN is well-measured, and shape and color cuts that restrict our sample to SNe Ia for which the SALT2 model is well-trained. Our cuts are summarized in Table 2 and Figure 2. Out of 3147 SNe with reliable host redshifts, SALT2 fits run successfully on

Table 1
Redshift Follow-up Summary

Telescope	Instrument	SN Redshifts ^a	$\lambda_{\min} - \lambda_{\max}$ Å	Avg. Exp. Time min.	Approx. Resolution Å pix ⁻¹	z_{median}
AAT	AAOmega	512	3700–8500	180	6	0.15
APO	DIS	10	3500–9800	60	2.5	0.24
MMT	Hectospec	2348	3700–9200	90	5	0.33
SDSS	BOSS	250	3800–9200	45	2.5	0.20
WIYN	Hydra	45	3700–6500	180	4.5	0.34
Other ^b	...	361	0.19
Total	...	3147	0.30

Notes. Some transient hosts were observed with multiple telescopes. Numbers include host galaxy observations of both spectroscopically confirmed and unconfirmed SN candidates.

^a Number of SN candidates with reliable redshifts.

^b Includes redshifts from 2dFGRS (Colless et al. 2003), 6dFGS (Jones et al. 2009), DEEP2 (Newman et al. 2013), VIPERS (Scodreggio et al. 2016), VVDS (Le Fèvre et al. 2005), WiggleZ (Blake et al. 2008), and zCOSMOS (Lilly et al. 2007).

Table 2
Sequential PS1 Data Cuts

	Removed	Remaining	This Cut Only	Without This Cut	Comments
Total candidates	...	5235
Host Sep. $R < 5$	774	4461	Likely host galaxy can be identified
Good host redshifts	1314	3147
Fit by SALT2	457	2690	SALT2 parameter fitting succeeds
Possible AGN	96	2594	2594	1040	Separated from center or no long-term variability
$-3.0 < X_1 < 3.0$	540	2054	2119	1092	SALT2 light-curve shape
$-0.3 < C < 0.3$	467	1587	1903	1215	SALT2 light-curve color
$\sigma_{\text{peakMJD}} < 2$	30	1557	2630	1021	Uncertainty in time of maximum light (rest frame days)
$\sigma_{X_1} < 1$	379	1178	1930	1386	X_1 uncertainty
Fit prob. ≥ 0.001	158	1020	2096	1178	χ^2 - and N_{def} -based probabilities from SALT2 fitter
$E(B-V)_{MW} > 0.15$	0	1020	2690	1020	Milky Way reddening

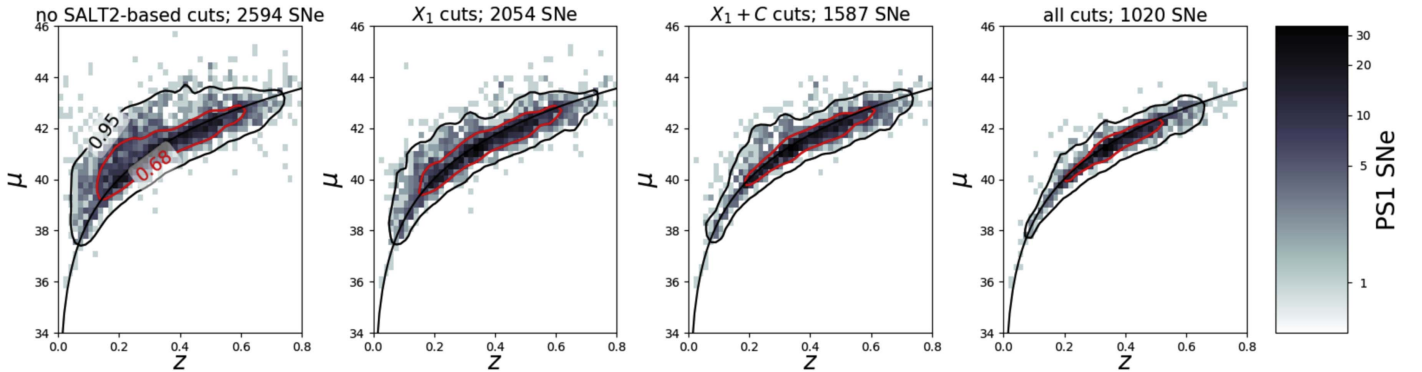


Figure 2. Effect of Betoule et al. (2014) cuts on the PS1 photometric Hubble diagram. Distance moduli are measured using the Tripp estimator (Equation (1)) with nuisance parameters from R14. Of the 2594 SNe that are fit by SALT2 and are not possible AGN shape and color cuts remove 1007, while χ^2 -based fit probability cuts and S/N-type cuts (shape uncertainty and time of maximum uncertainty) remove an additional 567 SNe, leaving 1020. Each set of cuts removes a mix of SNe Ia with poor light-curve quality and CC SNe.

2690 SNe (SALT2 parameter fitting often fails due to lack of light-curve data before or after maximum). A total of 1020 SNe pass all of our cuts.

Omitting the SALT2 σ_{X_1} cut has the largest single impact on our final sample. Without it, there would be nearly 1400 SNe in the sample but also twice as many SNe with Hubble residuals > 0.5 mag (poorly measured SNe Ia or CC SNe). The cut with the second largest reduction is the cut on C , without which there would be ~ 1200 SNe (though many would be CC SNe).

Although it may be possible to increase the SN sample size with relaxed cuts, the extent to which SNe Ia with low S/N and unusual colors are standardizable is not well-characterized.

In addition to the Betoule et al. (2014) cuts, we implement an additional set of cuts to remove possible AGN that were not flagged during the PS1 transient search. We tuned our long-term variability criteria to find known AGN in PS1 data. We found that sources where $> 25\%$ of background epochs have 2σ deviations from 0 are likely AGN (we define background

epochs as <20 days before or >60 days after the discovery epoch). Eighty-six SNe with both evidence of long-term variability and SN positions within $0''.5$ of their host centers were removed. After light-curve cuts, removing likely AGN reduces our sample by just 18 SNe. To have a sample with uniform selection, we make these cuts (and all cuts) regardless of whether or not a given SN Ia is spectroscopically confirmed.

2.3. Low- z SNe

Cosmological parameter constraints are greatly improved when a large, low- z SN Ia sample is included to anchor the Hubble diagram. We use the same 197 low- z SNe Ia used in R14 though we anticipate adding additional low- z SNe in our full cosmological analysis. These SNe are spectroscopically confirmed and are assumed to have no CC SN contamination.

The R14 PS1 cosmology analysis has a low- z sample with higher intrinsic dispersion than the PS1 sample. The intrinsic dispersion, σ_{int} , is defined as the value added in quadrature to the SN Ia distance modulus uncertainty such that the Hubble diagram reduced χ^2 is equal to 1 (Guy et al. 2007). Differences in SN Ia intrinsic dispersion from survey to survey are typical, with the likely source of the variation including underestimated photometric difference image uncertainties and excess scatter from bright host galaxy subtractions (as seen in R14 and Kessler et al. 2015). Redshift evolution of the SN Ia population could also play a role. We added 0.05 mag in quadrature to the m_B uncertainties of the low- z SNe to resolve the discrepancy. Once added, this additional uncertainty term gives both the PS1 and low- z SNe from R14 the same intrinsic dispersion of ~ 0.115 mag.

3. Simulating the Pan-STARRS Sample

To robustly determine how CC SN contamination affects PS1 measurements of w , we require a simulation that encapsulates as many elements of the PS1 SN survey as possible. We used the SuperNova ANALysis software (SNANA¹⁴; Kessler et al. 2009b) to generate Monte Carlo realizations of the PS1 survey. SNANA simulates a sample of SNe Ia and CC SNe using real observing conditions, host galaxy noise, selection effects, SN rates, and incorrect redshifts from host galaxy mismatches or measurement error. Simulations assume a flat Λ CDM cosmology with $H_0 = 70 \text{ km s}^{-1} \text{ Mpc}^{-1}$, $\Omega_M = 0.3$, $\Omega_\Lambda = 0.7$, and $w = -1$.

We choose not to simulate one significant effect: the correlation between SN luminosity and host mass (the host mass bias; Kelly et al. 2010; Lampeitl et al. 2010). We do not simulate the host mass bias because R14 did not include it (finding it had low significance in their sample), and we wish to compare our PS1 photometric results directly to those of R14. This effect has been identified at $>5\sigma$ by Betoule et al. (2014), and we will include it in our future cosmological analysis with these data.

Each major component of our simulation is discussed in detail below.

1. *Observing conditions.* SNANA generates SN observations based on a simulation library file with observation dates, filters, sky noise, zeropoints, and PSF sizes that we measure from PS1 nightly images.

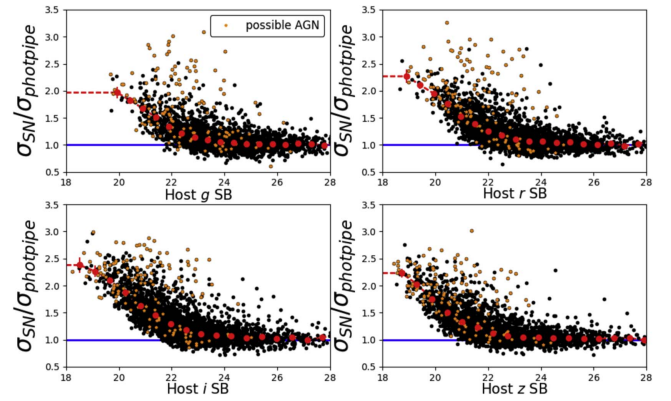


Figure 3. Ratio between “true” and DAOPHOT-derived photometric noise as a function of host galaxy surface brightness in the $griz_{PS1}$ filters. We computed the host galaxy surface brightness by averaging over one PSF FWHM at the SN location. We computed the true photometric noise by inflating the errors from DAOPHOT (which do not include host galaxy noise) such that light-curve epochs without SN light had $\chi^2 = 1$. Possible AGN (gold stars) comprise many of the outliers in this relation. We incorporated this relationship into our SNANA simulations to yield an accurate prediction of photometric uncertainties.

2. *Host galaxies.* The observed flux scatter of SNe found in bright galaxies exceeds what is expected from Poisson noise alone (R14; Kessler et al. 2015). To correct for this, SNANA adds host galaxy noise to SN flux uncertainties by placing each SN in a simulated host galaxy. The SN is placed at a random location that has been weighted by the galaxy surface brightness profile. The distribution of PS1 host galaxies was determined from PS1 data; we measured the magnitudes and shape parameters of PS1 SN host galaxies using SExtractor, with zeropoints measured from the PS1 pipeline. We then use the noise model from Kessler et al. (2015, their Equation (4)):

$$\tilde{\sigma}_{\text{flux}} = \sigma_{\text{flux}} \times R_\sigma, \quad (2)$$

where R_σ is a function of host galaxy surface brightness (the vertical axis of Figure 3). We determine R_σ for PS1 by comparing host surface brightness to the flux error scaling that gives light-curve epochs without SN flux a reduced $\chi^2 = 1$.

3. *Selection effects.* Two primary selection effects come into play in a photometric SN Ia survey. The first is detection efficiency, the fraction of single-epoch detections as a function of the photometric S/N. The detection efficiency is computed by dividing the number of epochs detected by PS1 at a given S/N by the total number of epochs at that S/N. SNANA uses the efficiency versus S/N, measured by PS1, to determine which simulated epochs are detected. SNANA then applies the PS1 survey requirement of three detections to “discover” an SN. The PS1 detection efficiency is $\sim 50\%$ for epochs with an S/N of 5 in the final light curves.

The second effect is host galaxy redshift selection. To model this effect, we incorporated a redshift-dependent “host galaxy efficiency” distribution in our simulations, which we adjusted such that the redshift distribution of the simulations matched our data.

4. *Uncertainty adjustment.* SNANA allows its simulated uncertainties to be scaled as a function of S/N such that the mean uncertainties in simulations match the mean uncertainties of our data. In PS1, this requires a modest

¹⁴ <http://snana.uchicago.edu/>

~5%–10% noise increase at low S/N (after excess host galaxy noise is added). This adjustment is necessary due to the non-Gaussian wings of the PS1 PSF and the PSF fitting radius used by the PS1 pipeline.

5. *Mismatched host galaxies and incorrect redshifts.* As discussed in Section 2.1, we expect $2.6\% \pm 1.4\%$ of our redshifts to be incorrect due to mismatched host galaxies and redshift measurement uncertainties. We used SNANA to simulate incorrect host redshifts by assigning false, “measured” redshifts to 2.6% of our SNe. These redshifts are drawn from a flat, random distribution between $z=0.01$ and $z=0.75$. This is the range of redshifts at which PS1 can discover SNe, with the exception of rare superluminous SNe. Superluminous SNe typically have hosts too faint for our follow-up survey to measure their redshifts (Lunnan et al. 2015).

We find that ~50% of SNe with incorrect redshifts fail our sample cuts, giving a final contamination fraction of ~1%–1.5%. In large part, this reduction is due to cuts on the SALT2 color parameter. If an SN has an incorrect redshift, SALT2 is twice as likely to infer that its observed-frame colors are inconsistent with normal SNe Ia when transformed to the wrong rest frame.

6. *SN Ia model.* The SN Ia model used in these simulations is the Guy et al. (2010) model with SN Ia nuisance parameters from R14 (SALT2 $\alpha = 0.147$, $\beta = 3.13$). The parent X_1 and C distributions were determined by Scolnic & Kessler (2016) for the PS1 spectroscopic sample. We adjusted the parent means of the X_1 and C distributions by 1σ to better match our data, making X_1 lower by 0.17 and C higher by 0.023. This difference is likely physical; on average, X_1 is lower and C is higher in massive host galaxies (e.g., Childress et al. 2013). Our host follow-up program preferentially obtained redshifts of massive galaxies.
7. *CC SN templates and diversity.* CC SNe are simulated based on a library of 43 templates in SNANA. The templates we use were originally created for the SN Photometric Classification Challenge (Kessler et al. 2010) and also used by Bernstein et al. (2012). Templates are based on bright, spectroscopically confirmed SDSS, SuperNova Legacy Survey (SNLS; Conley et al. 2011; Sullivan et al. 2011), and Carnegie Supernova Project (Hamuy et al. 2006; Stritzinger et al. 2011) CC SNe with well-sampled light curves. Templates were created from the light curves by warping a model spectrum for each SN subtype to match the light-curve fluxes in every broadband filter (see Appendix A.1.1).

SNANA has 24 II-P templates, 2 II-n templates, 1 II-L template, 7 Ib templates, and 9 Ic templates. In this work, we make the assumption that reddening in the templates is approximately equal to reddening in our data. This assumption allows us to use the Li et al. (2011) LFs, which have not been corrected for reddening, and SNANA templates, which also include intrinsic reddening. Correcting these templates, the Li et al. (2011) rates and LFs for reddening are an important avenue for future work.

We added a subtype-specific magnitude offset to each CC SN template such that the mean simulated absolute magnitude of the subtype matched the mean of its Li et al. (2011) LF. By applying a uniform offset to every template in a subtype, the brightness of different templates relative

to their subtype is incorporated in our simulations.¹⁵ We also matched the dispersions of the Li et al. (2011) LFs by adding an additional, random magnitude offset to each simulated CC SN. This offset was drawn from a Gaussian with a width we adjusted such that the dispersion of the simulated absolute magnitudes for each subtype matched that of Li et al. (2011).

8. *SN Rates.* SNANA creates a combined SN Ia+CC simulation, with each SN type normalized by its rate. The redshift-dependent SN rates used in this work are the same as the baseline model of Rodney et al. (2014). SNe Ia follow measured rates, while CC SNe follow the cosmic star formation history. Relative rates of SN types and subtypes are anchored at $z=0$ by Li et al. (2011) and evolve $\propto(1+z)^\gamma$, where γ is a free parameter tuned to match theory and observations (only a single value for γ is needed over the redshift range of PS1). We used $\gamma_{\text{Ia}} = 2.15$ and $\gamma_{\text{CC}} = 4.5$ (Rodney et al. 2014).

Figure 4 compares our simulations to the data after fitting all SNe with the SALT2 model. Note that CC SN information in this simulation is obtained without any PS1 analysis or input. SALT2 fitting is an effective way to examine both SNe Ia and the light-curve parameters of Ia-like CC SNe. Discrepancies in Figure 4 indicate potential biases when measuring cosmological parameters with a CC SN-contaminated sample.

Our simulations agree closely with the data for most light-curve parameters. The maximum S/N of the simulated light curves matches the data (Figure 4(B)), as does the distribution of SALT2 X_1 (Figure 4(C)). However, there are too few simulated SNe with red SALT2 colors (Figure 4(D)). The simulated redshift evolution of X_1 and C matches the data well (Figures 4(F) and (G)).

Though most simulated light-curve parameters match our data well, the Hubble residuals (Figure 4(A)) show a discrepancy. We see ~3 times more SNe than expected between $0.5 \lesssim \mu - \mu_{\Lambda\text{CDM}} < 1.5$ mag (these SNe are fainter than SNe Ia at their redshifts). For this reason, we used light-curve-based classifications of our data to adjust the CC SN LFs. The details of this procedure are discussed in Appendix A. We find that the peak of the CC SN LF must be brightened by 1.2 mag for SNe Ib/c and 1.1 mag for SNe II in order for our simulations to match our data (Figure 5). The dispersion of CC SN templates must be reduced by 55% for SNe Ib/c. We also add four 1991bg-like SN Ia templates and four SNI Ib template to SNANA to include a broader range of SN types.

The CC SNe LFs in our adjusted simulation are ~5 σ brighter than in Li et al. (2011). However, these results do not necessarily imply that the true LFs of CC SNe show a ~5 σ inconsistency with Li et al. (2011). Rather, they indicate that our SALT2-based shape and color cuts isolate a region of CC SN parameter space that is not the average. Although we find it plausible that the CC SNe with shapes and colors most similar to SNe Ia have brighter and lower-dispersion LFs than CC SNe as a whole, further work is required to understand the diversity of CC SN subpopulations. Larger sets of high-cadence, high-quality spectral time series from which to construct templates are also necessary. An additional factor is that the low statistics in the LOSS volume-limited sample

¹⁵ We tweaked this procedure for SNe Ib, which had one anomalously bright template. All SN Ib templates were adjusted by individual magnitude offsets such that each template matched the mean magnitude of SNe Ib given by Li et al. (2011).

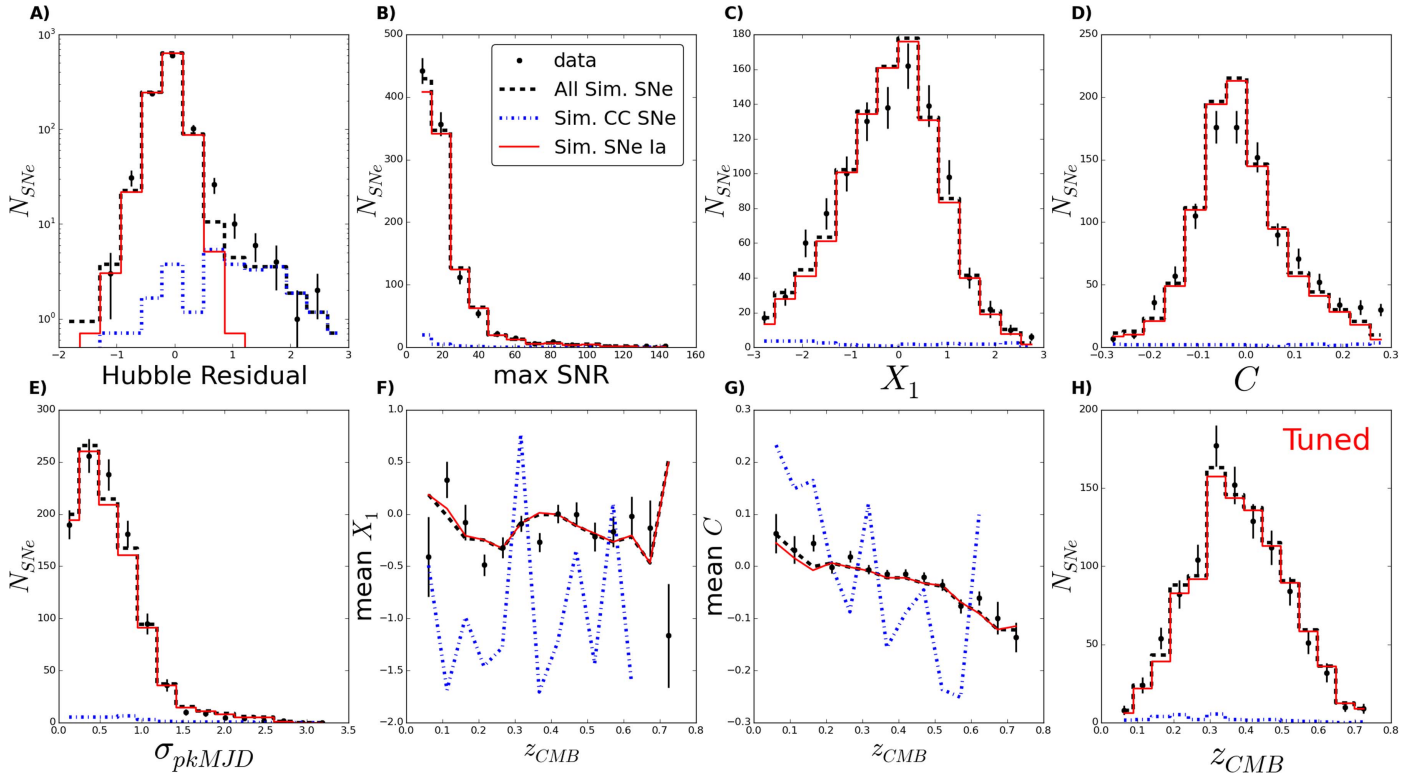


Figure 4. SNANA simulations of a PS1 photometric sample compared to PS1 data. The simulated Hubble residuals (A) of the CC SN distribution are flatter and fainter than the data. The simulated S/N (B), shapes (C), colors (D), uncertainties (E), and X_1/C redshift dependencies (F and G) match our data closely, albeit with $\sim 3\sigma$ discrepancies in the time of maximum uncertainty and SN color. We tuned the simulated redshift distribution (H) to match our data.

require that the shape of the CC SN LFs be extrapolated in some way. We treat CC SN LFs as Gaussians, which is most likely a flawed assumption (see Figure 16 of Li et al. 2011).

4. Estimating SN Ia Distances with BEAMS

We use the BEAMS method to obtain SN Ia distance measurements that are corrected for the CC SNe contaminating our data (KBH07). The implementation of BEAMS suggested in KBH07 solved for distances and cosmological parameters in a single step; here, we first use BEAMS to solve for binned SN Ia distances and then use CosmoMC (Lewis & Bridle 2002) to determine cosmological parameters. This procedure will allow us to more easily combine SN data with complementary CMB and BAO data in our forthcoming cosmological analysis. We summarize the method below.

BEAMS simultaneously determines Ia and CC SN distances by sampling a posterior probability distribution that includes both SN Ia and CC SN populations in the likelihood. The BEAMS posterior, the probability of the free parameters θ given the data, D , is proportional to the product of the individual likelihoods for each SN multiplied by the priors on the free parameters:

$$P(\theta|D) \propto P(\theta) \times \prod_{i=1}^N \mathcal{L}_i. \quad (3)$$

The simplest suggested likelihood from KBH07 uses Gaussian distributions to represent CC SN and SN Ia

populations:

$$\begin{aligned} \mathcal{L}_i = & P_i(\text{Ia}) \times \frac{1}{\sqrt{2\pi(\sigma_{i,\text{Ia}}^2 + \sigma_{\text{Ia}}^2)}} \exp\left(-\frac{(\mu_{i,\text{Ia}} - \mu_{\text{Ia}}(z_i))^2}{2(\sigma_{i,\text{Ia}}^2 + \sigma_{\text{Ia}}^2)}\right) \\ & + P_i(\text{CC}) \times \frac{1}{\sqrt{2\pi(\sigma_{i,\text{CC}}^2 + \sigma_{\text{CC}}(z_i)^2)}} \\ & \times \exp\left(-\frac{(\mu_{i,\text{CC}} - \mu_{\text{CC}}(z_i))^2}{2(\sigma_{i,\text{CC}}^2 + \sigma_{\text{CC}}(z_i)^2)}\right). \end{aligned} \quad (4)$$

$P_i(\text{Ia})$ is the prior probability that the i th SN is of Type Ia. $P_i(\text{CC})$, the probability that the SN is a CC SN, is equal to $1 - P_i(\text{Ia})$. $\mu_{i,\text{Ia}}$, $\mu_{i,\text{CC}}$ and $\sigma_{i,\text{Ia}}$, $\sigma_{i,\text{CC}}$ are the distance modulus and distance modulus uncertainties for the i th SN, derived using the Tripp estimator (Equation (1)). We differentiate between measured Ia and CC distance moduli from the data because we will allow the Tripp estimator to use different nuisance parameters for the SN Ia and CC SN terms in the likelihood (Section 4.1). μ_{Ia} , σ_{Ia} and μ_{CC} , σ_{CC} are the means and standard deviations of the SN Ia and CC SN Gaussians, respectively.

The variables μ_{Ia} and μ_{CC} are a function of the redshift, z , of the i th SN and of cosmological parameters. The variable σ_{CC} is redshift dependent as well, primarily due to the changing mix of CC SN subtypes that PS1 is able to discover as a function of redshift. We fit for $\mu_{\text{Ia}}(z)$, $\mu_{\text{CC}}(z)$, and $\sigma_{\text{CC}}(z)$ by allowing BEAMS to treat them as free parameters at certain fixed

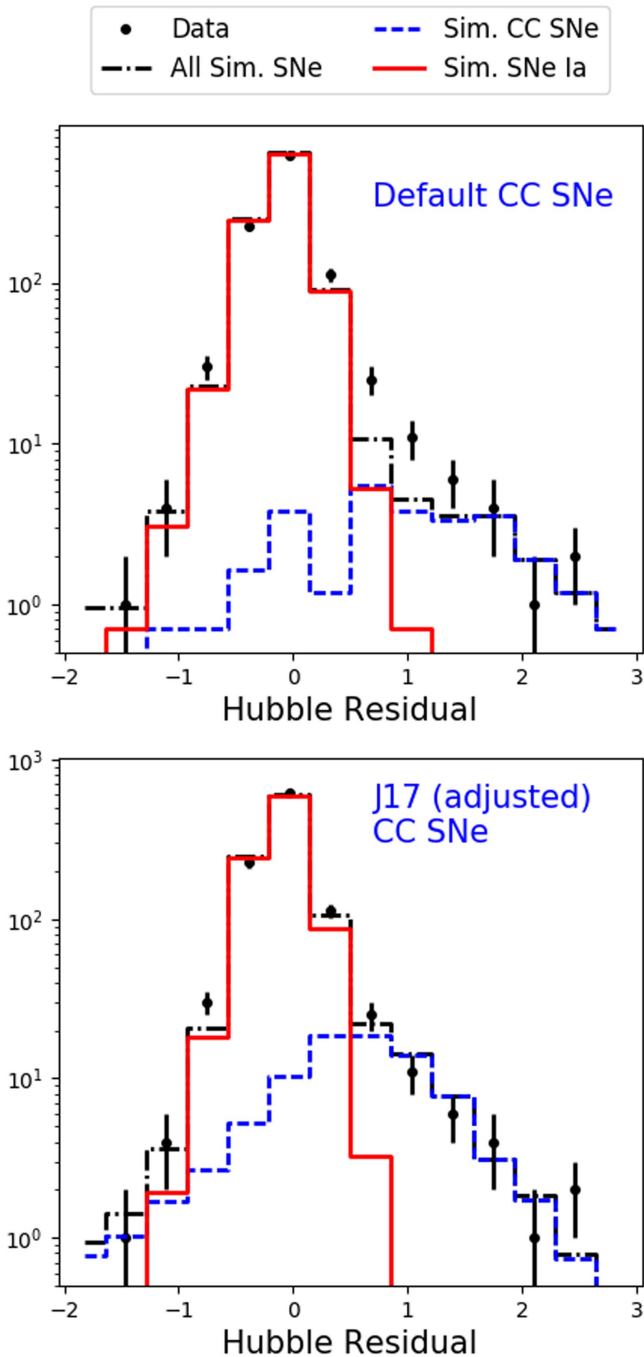


Figure 5. Comparison of Hubble residuals before and after empirical adjustments to CC SN LFs. We enlarge Figure 4(A) (top) and compare to our adjusted J17 simulations (bottom). Before empirical adjustments, the simulations contained just 2.4% CC SNe and were a poor match to the data. After adjustments, the simulations have 8.9% CC SNe. Discrepancies between data and simulations in the red end of the SALT2 C distribution can be explained by additional CC SNe.

redshifts z_b . We refer to the set of fixed redshifts as “control points” following Betoule et al. (2014).¹⁶ Between two control points, the distance modulus (and dispersion) is interpolated by

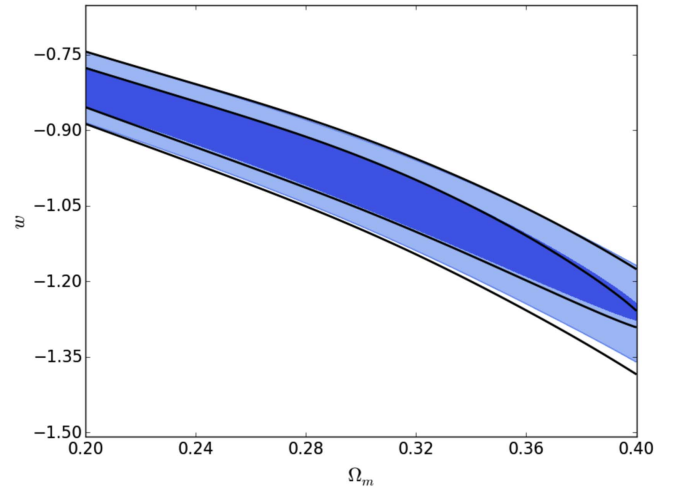


Figure 6. Comparing the full SN Ia likelihood (filled contours) to the binned Ia likelihood (black).

a linear function of $\log(z)$ defined by

$$\mu(z) = (1 - \xi)\mu_b + \xi\mu_{b+1} \quad (5)$$

$$\xi = \log(z/z_b) / \log(z_{b+1}/z_b),$$

where μ_b is the distance modulus at redshift z_b .

Betoule et al. (2014) fit to a set of 30 log-spaced redshift control points and found that the difference between Λ CDM and the interpolation is always smaller than 1 mmag. We used 25 control points for the smaller PS1 redshift range of $0.01 < z < 0.7$ (we restrict our sample to $z < 0.7$, as very few PS1 SNe can be found at higher redshifts). In Figure 6, we compare the cosmological constraints from 1000 individual SNe Ia to the approximate results derived from the SN Ia distances at 25 control points ($P(\text{Ia}) = 1$ for all SNe Ia). We find that the cosmological constraints are nearly identical.

We use five log-spaced redshift control points for CC SNe. If true SN type probabilities are known, five CC SN control points allows BEAMS enough flexibility to avoid biasing the Ia likelihood with a poor determination of the CC SN distribution. We allow the intrinsic width of the CC SN Gaussian distribution (σ_{CC}) to vary with redshift, but keeping the intrinsic width of the SN Ia Gaussian fixed. By using the SALT2mu procedure (Marriner et al. 2011), we verified that the (simulated) uncertainty-weighted dispersion of SNe Ia does not change with redshift for PS1 (this is also a typical assumption in cosmological analyses; Guy et al. 2010). This physically realistic assumption gives BEAMS more leverage to discriminate between SNe Ia and CC SNe, which have much higher dispersion than SNe Ia.

In total, our baseline implementation of BEAMS has 38 free parameters: 25 SN Ia distance moduli at Ia control points, 5 CC SN distance moduli at CC control points, 5 CC SN dispersion parameters, 1 SN Ia dispersion parameter,¹⁷ and the SALT2 nuisance parameters α and β which are used to compute μ_i and σ_i (discussed below). BEAMS free parameters can be efficiently estimated by sampling the logarithm of the posterior with a Markov Chain Monte Carlo (MCMC) algorithm. This work uses emcee,¹⁸ a Python MCMC

¹⁶ Note that Betoule et al. (2014) use this method to increase computational efficiency when combining SN Ia data with Planck priors. However, their method of reducing SN data to a set of distances at redshift control points is well-suited for a BEAMS-like algorithm.

¹⁷ Throughout, we have written this dispersion parameter as σ_{Ia} to distinguish it from σ_{int} , the global uncertainty term used in many previous analyses. σ_{int} , defined in Section 2.3, has a different definition from the BEAMS free parameter σ_{Ia} .

¹⁸ <http://dan.iel.fm/emcee/current/>

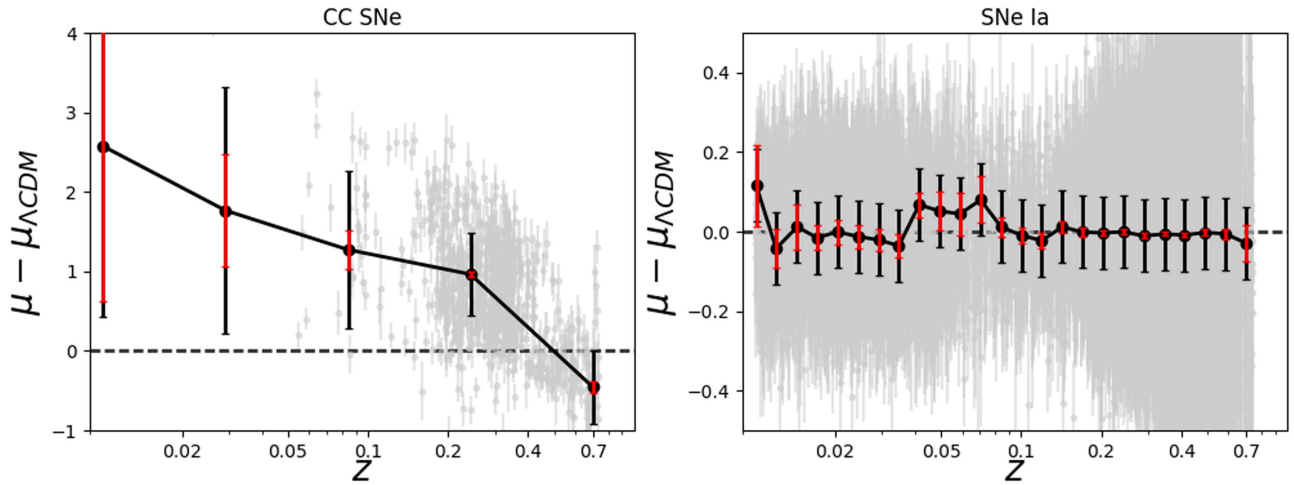


Figure 7. Illustration of BEAMS. Simulated CC SNe (left) and SNe Ia (right) with the redshift-dependent BEAMS parameters μ_{CC} , μ_{Ia} (black points) and σ_{CC} , σ_{Ia} (black bars). Uncertainties on μ_{CC} and μ_{Ia} are in red. We use correct prior probabilities of $P(Ia) = 1$ for SNe Ia with correct redshifts and $P(Ia) = 0$ for all others.

implementation (Foreman-Mackey et al. 2013). We use `emcee`'s Parallel-Tempered Ensemble Sampler to explore the multimodal peaks of the likelihood robustly. Figure 7 illustrates the Hubble residual diagram from BEAMS using simulated SNe and correct prior probabilities (all SNe Ia with correct redshifts have $P(Ia) = 1$ and all other SNe have $P(Ia) = 0$). Note that if few or no CC SNe are in a given redshift bin, the magnitude and uncertainty of CC SN distances are primarily determined by the priors.

We apply loose Gaussian priors on most BEAMS free parameters, but find that with samples of 1000 SNe or more, the difference between Gaussian and flat priors is negligible. For SN Ia distances, we apply flat priors. Though we assume some prior knowledge of the CC SN distribution, our priors on CC SN distance (μ_{CC} in Equation (4)) are very loose; we use broad Gaussians of width 3 mag that are centered at 2 mag fainter than the SNe Ia at each control point. SALT2 nuisance parameters have Gaussian priors of width five times the uncertainties from R14. Our code is available online.¹⁹

4.1. SALT2 Light-curve Parameters

We use a SALT2 fitting program to measure SN light-curve parameters for our sample. However, SALT2 parameters do not directly measure the distance modulus (Equation (1)). For BEAMS to measure distances using SALT2 light-curve fits, the nuisance parameters α and β must either be fixed to the value from a spectroscopic sample or incorporated into BEAMS as free parameters. We allow α and β to be free parameters here as it is a more general test of the method. Different survey methods, detection efficiencies, and selection criteria can significantly bias recovered SN parameters (Scolnic & Kessler 2016), which could make it necessary for future analyses to be able to fit for these parameters. In the CC SN component of the BEAMS likelihood, we fixed α and β to the nominal value for SN Ia spectroscopic samples (allowing them to float has no effect on our results).

Because we include α and β as free parameters, the likelihood presented in Equation (4) has a term in the Gaussian normalization factor, σ_i , that depends on α and β . The result is a significant bias in the derived SN parameters (March

et al. 2011). This bias grows for larger SN samples (see Appendix B of Conley et al. 2011 and Kelly 2007 for details). The solution adopted in Conley et al. (2011) is to neglect the normalization term when determining α , β , and σ_{int} by using a simple likelihood $\mathcal{L} \propto \exp(-\chi^2/2)$. For 1000 SNe, Conley et al. (2011) find that the bias from this likelihood is well below the statistical error. Though we cannot use this solution without biasing determinations of the CC SN and SN Ia distributions, we use an alternative formalism and treat the uncertainties on the distance modulus as fixed in the denominator of the normalization term (independent of α and β). Fixing distance modulus uncertainties in the denominator does not bias α , β , or w and is a very modest approximation; in the PS1 sample, varying α and β within their 1σ errors from R14 gives a mean change in uncertainty of only 2 mmag. No individual SN has its uncertainty change by >20 mmag. See Kessler & Scolnic (2017, Section 8.1) for an alternative solution.

4.2. Prior Probabilities

The BEAMS formalism requires an estimate of the prior probability that a given SN is of Type Ia. This prior can be measured by an SN classifier or it can be as simple as setting $P(Ia) = 1/2$ for all SNe. For our baseline analysis, we adopt the PSNID light-curve fitter, as implemented in SNANA (Sako et al. 2011, 2014). In PSNID, observed SN light curves are fit with perfect, noise-free simulations of the SALT2 SN Ia model and SNANA's CC SN templates to determine the probability that each SN is of Type Ia.²⁰ PSNID estimates $P(Ia)$ from the χ^2 of the fit and includes type, redshift, and luminosity priors. The set of SNe with $P(Ia) > 0.5$ has 2.9% contamination by CC SNe while including 92% of real SNe Ia.

We allow a remapping of the PSNID prior probabilities by adding two parameters to BEAMS: one that re-normalizes the probabilities and a second that shifts them linearly. The first parameter is a scaling factor that corrects for globally skewed prior probabilities following Hlozek et al. (2012). This normalization term allows BEAMS to correct for effects such

¹⁹ See Jones (2017), with recent updates at <https://github.com/djones1040/BEAMS>. Example input files are also provided.

²⁰ Because the simulated CC SN models in SNANA are the same as the CC SN models in the PSNID template library, we used an option in PSNID (SNANA v10_47m and later) that ensures a CC SN simulated using a given template cannot be classified using a noise-free version of that same template. This option increases the CC SN contamination by $\sim 1\%$.

as incorrect redshift-dependent SN rates, inaccurate classifier training, or other $P(\text{Ia})$ biases. The second parameter is a global, linear shift in probability to handle incorrect typing near $P(\text{Ia}) = 0$ or $P(\text{Ia}) = 1$ (but requiring $0 < P(\text{Ia}) < 1$). This is necessary in cases where uncertainty in $P(\text{Ia}) \simeq 1$ or $P(\text{Ia}) \simeq 0$ is significant (KBH07). The relationship between the normalization factor, A , the shift parameter, S , and the probability $P(\text{Ia})$ is given by

$$\tilde{P}(\text{Ia}) = \frac{A \times (P(\text{Ia}) + S)}{1 - (P(\text{Ia}) + S) + A \times (P(\text{Ia}) + S)} \times 0 < \tilde{P}(\text{Ia}) < 1. \quad (6)$$

Another solution suggested by KBH07 that could be explored in future work is adding a probability uncertainty term to the likelihood.

4.3. Malmquist Bias

$\mu_{i,\text{Ia}}$, the SALT2-derived distance modulus for the i th SN, is subject to Malmquist bias for magnitude-limited surveys such as PS1. We account for the SN Ia Malmquist bias using PS1 and low- z simulations to determine the redshift-dependent bias of derived SN Ia distances. We used Monte Carlo simulations of $\gtrsim 10,000$ SNe and non-parametric spatial averaging to determine and correct for the trend in distance modulus. Our spatial averaging algorithm uses local polynomial smoothing to interpolate the mean distance modulus trend across the redshift range.

Our simulations of the spectroscopically confirmed low- z SN sample follow R14, who use the same α and β as our PS1 simulations. The details of these low- z simulations and the determination of the spectroscopic selection function are discussed in detail in Scolnic et al. (2014a, see their Figure 6 for a comparison between simulations and data).

Figure 8 shows the simulated, redshift-dependent measurement bias in distance modulus, m_B , αX_1 , and $-\beta C$. The average high- z distance modulus bias in PS1 is nearly identical to the bias measured for PS1 spectroscopically confirmed SNe by R14. One difference is that the Malmquist bias is almost negligible in our sample until $z \sim 0.35$. Some differences in bias are expected because the R14 bias is dominated by their spectroscopic SN follow-up selection function.

At $z > 0.5$, we find that the bias in X_1 , C , and μ becomes large as flux uncertainties near the epoch of peak brightness are up to a factor of 2 larger than in the lower- z data. Greater than 50% of the m_B and C bias at these redshifts is due to our cut on X_1 uncertainty, which is effectively an S/N cut that increases the selection bias. Distance biases due to cuts on X_1 and C are also expected as the data become noisier and statistical fluctuations cause more SNe that fall outside the luminosity-correlated range to appear on our Hubble diagram (Scolnic & Kessler 2016, see their Figure 4). Our simulations also show that requiring lower X_1 uncertainty tends to select narrower measured light-curve shapes. Accordingly, Figure 4(F) shows that the measured X_1 distribution remains largely flat with redshift; although SNe with larger X_1 values are intrinsically more luminous and thus more likely to be discovered, the measurement bias shown in Figure 8 has an opposite, and approximately equal, effect.

A discussion of systematic error in Malmquist bias determination will be presented in our forthcoming cosmological analysis. This will include incorporating α and β

uncertainties, which can cause differences in the distance bias of ~ 5 mmag at $z > 0.5$. Although Scolnic et al. (2014a) found that the Malmquist bias is not one of the dominant sources of error, the photometric sample may be subject to different biases than a typical spectroscopic sample due to its lower average S/N.

We correct all SNe, but only for the SN Ia Malmquist bias (we do not attempt bias corrections based on $P(\text{Ia})$). It is not necessary to correct for the CC SN Malmquist bias, as CC SNe are not used to derive cosmological parameters. However, we implicitly model the CC SN Malmquist bias using BEAMS because BEAMS allows the CC SN mean and dispersion to vary with redshift.

4.4. Cosmological Parameter Fitting

Finally, once distance moduli at the 25 redshift control points have been measured with BEAMS, BEAMS distances and distance covariance matrices can be used as inputs into the Cosmological Monte Carlo software for cosmological parameter fitting (CosmoMC; Lewis & Bridle 2002). For computational efficiency, we did not use the full Planck chains in this analysis and instead ran CosmoMC on our BEAMS results with a Planck-like prior of $\Omega_M = 0.30 \pm 0.02$.

5. Cosmological Results from BEAMS

5.1. Tests with Simulated Data

We generated 25 simulations of 1000 PS1 SNe each (25,000 total SNe) in order to test BEAMS on samples the size of the PS1 photometric sample. We add simulated low- z samples of 250 SNe Ia each, the approximate number that will be included in our forthcoming cosmological analysis. The results presented here use the J17 CC SN simulations (Appendix A), as they have CC SN LFs that match our data.

To focus on biases from CC SN contamination, we define the CC SN bias Δ and the increase in statistical uncertainty due to CC SNe, $\Delta\sigma_{\text{stat}}$, for a given parameter P :

$$\Delta = P_m - P_{\text{Ia}}, \quad \Delta\sigma_{\text{stat}} = \sigma(P_m) - \sigma(P_{\text{Ia}}), \quad (7)$$

where P_m is the measured parameter from the BEAMS method and P_{Ia} is the measured parameter from the BEAMS method using SNe Ia alone and setting all prior probabilities equal to one. For the 25 simulated samples, the average w_{Ia} value is -1.001 ± 0.009 . The rms of w_{Ia} is 0.045, consistent with the mean statistical uncertainty (0.048).

We compare the Ia-only distances, SN parameters, and w measurements against our results from the BEAMS method in Table 3. Figure 9 shows that the binned distances are biased by less than 20% of their uncertainties with the exception of the final control point. Typical biases are ~ 3 mmag and the largest average bias from the 25 samples (aside from the final high-uncertainty control point) is 6 mmag at $z \simeq 0.6$.

The SN parameters α and β are biased by 3%, or 1–1.5 times the average statistical error. σ_{Ia} is biased by 4%, 0.3 times the average statistical error. Note that σ_{Ia} (in Equation (4)) is functionally similar to the SN intrinsic dispersion, σ_{int} . These biases are small enough that they would be difficult to measure in real data. A possible cause of these biases is that Ia-like CC SNe have color laws more consistent with Milky Way dust ($\beta \sim 4.1$) and different shape–luminosity correlations.

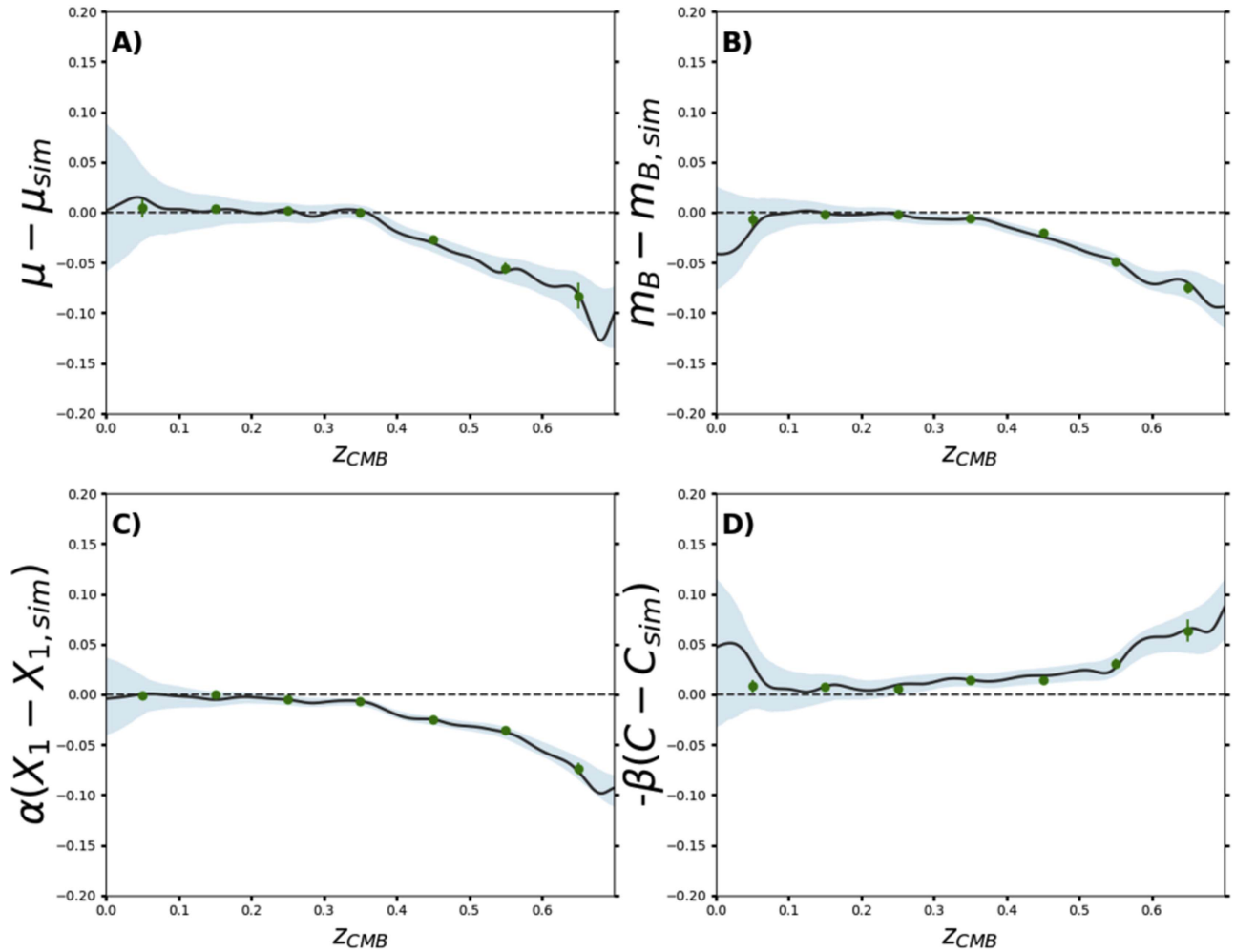


Figure 8. Simulated redshift-dependent bias in (A) distance, (B) peak B magnitude, (C) αX_1 , (D) and $-\beta C$ for the PS1 photometric sample using non-parametric spatial averaging (black lines with 95% confidence intervals in blue) with median bins (points) shown for comparison. The PS1 sample has negligible distance (Malmquist) bias until $z \sim 0.3$ and a maximum bias of ~ 0.1 mag at $z \gtrsim 0.6$.

We find that w has a median bias of -0.005 ± 0.004 due to CC SN contamination, 10% of the statistical error on w . While our analysis is consistent with no bias, we assign a systematic uncertainty on w of $0.005 + 0.004 = 0.009$, though the true systematic uncertainty could be higher due to uncertainties in CC SN simulations (Section 6). The statistical uncertainty on w in this case is just 3% higher than the statistical uncertainty from SNe Ia alone. This result is consistent with KBH07, who find that BEAMS can yield nearly optimal uncertainties (we discuss BEAMS uncertainties further in Section 7.3).

If we compare the bias on w to a naïve method of measuring w with photometrically classified SNe, the advantage of using BEAMS is obvious. For our twenty-five 1000 SN samples, we take likely SNe Ia ($P_{\text{PSNID}}(\text{Ia}) > 0.5$) and estimate cosmological parameters assuming that all of these SNe are Type Ia (Campbell et al. 2013 used a similar method of cutting the sample based on PSNID classifications). Making this cut removes 8% of the true SNe Ia in our sample and yields a final sample contaminated by 2.9% CC SNe. In spite of having a sample comprised of $>97\%$ SNe Ia, the average bias on w is -0.025 ± 0.004 , a factor of five higher than our BEAMS results. The bias is $>50\%$ of the statistical uncertainty on w and has 6σ significance, while the BEAMS result is consistent with no bias. The statistical uncertainty on w from this method is 6%

higher, compared to 3% higher from BEAMS. Even a cut of $P_{\text{PSNID}}(\text{Ia}) > 0.9$ yields a bias on w of 0.011 ± 0.003 ($>3\sigma$ significance) at the cost of removing 17% of real SNe Ia. Furthermore, while BEAMS allows these probabilities to be adjusted by the method, treating them as fixed in this simplistic method increases the possibility of biased classifications due to incompleteness in the CC SN template library. It is clear that BEAMS outperforms this simple cut-based analysis, though this naïve method could still be effective with significantly improved classification methods.

5.2. Comparing Real Pan-STARRS Photometric Supernovae to Rest et al. (2014)

Rather than analyzing the full PS1 sample, we analyze 25 random draws of PS1 SNe to compare R14 measurements—and uncertainties—directly to measurements from CC SN-contaminated samples of the same size. Because 96 R14 SNe Ia pass our sample cuts, we draw samples of 104 photometric SNe in order that our subsamples each contain an average of 96 SNe Ia (and 8 CC SNe; we also use reprocessed R14 light curves). We do not explicitly require these random samples to have the same redshift distribution as the PS1 spectroscopic sample. However, the redshift distribution of the PS1 photometric sample is similar to that of R14 (a nearly identical

Table 3
Results from BEAMS

	PS1 Simulations					PS1 Data				
	Bias	$\sigma_{\text{bias}}^{\text{a}}$	$\sigma_{\text{stat}}^{\text{b}}$	bias/ σ_{stat}	$\Delta\sigma_{\text{stat}}$	Bias	$\sigma_{\text{bias}}^{\text{a,c}}$	$\sigma_{\text{stat}}^{\text{b}}$	bias/ σ_{stat}	$\Delta\sigma_{\text{stat}}$
μ^{d}	0.000	0.001	0.031	0.0	0.001 (3%)	−0.040	0.019(±0.09)	0.074	−0.4	0.010 (14%)
α	0.004	0.000	0.006	0.6	0.000 (3%)	0.001	0.001(±0.005)	0.012	0.1	0.001 (5%)
β	0.088	0.008	0.073	1.2	0.004 (6%)	0.199	0.018(±0.10)	0.154	1.4	0.009 (6%)
w	−0.005	0.004	0.048	−0.1	0.002 (3%)	−0.040	0.012(±0.084)	0.095	−0.4	0.008 (8%)

Notes. Bias and increase in uncertainty due to CC SN contamination. All quantities shown are taken from the median of 25 samples. Bias is defined in Equation (7) for simulations and Equation (8) for data (bias in data is relative to R14 parameter measurements).

^a Uncertainty on the median bias.

^b Statistical uncertainty on each parameter from a single sample.

^c In parentheses, we show the estimated uncertainty on the R14 values. Because our PS1 data are correlated with R14 (they share the low- z sample), we take Monte Carlo samples of 100 simulated PS1 SNe and combine each with the R14 low- z sample, taking the standard deviation of measurements from these combined data as the uncertainty.

^d Averaged over $0.08 < z < 0.7$.

range and median redshift, though the photometric sample does include more faint SNe Ia with red colors).

For subsamples of PS1 data, we report parameter biases relative to R14:

$$\begin{aligned}\Delta &= P_m - P_{\text{R14}} \\ \Delta\sigma_{\text{stat}} &= \sigma(P_m) - \sigma(P_{\text{R14}}),\end{aligned}\quad (8)$$

where P_{R14} and $\sigma(P_{\text{R14}})$ refer to a parameter and its uncertainties from R14.

Although R14 does not have enough SNe to test for small biases in w , the data still allow for a consistency check that is independent of the myriad assumptions made in simulations. In addition, the 96 SNe from R14 with low- z SNe can provide constraints on the bias of the nuisance parameters α , β , and σ_{Ia} due to the BEAMS method. We include low- z SNe because BEAMS is more robust when it has a spectroscopically confirmed sample as part of the data and has difficulty measuring accurate SN Ia dispersions for small samples.

We find that the measured distances, SN nuisance parameters α and β , and w are consistent with R14 (Table 3). We may be seeing the same hints of a bias toward higher values of β that we find in simulations but they have under 2σ significance. The bias in α is not statistically significant (0.1σ).

The average of w from the twenty-five 104 SN samples is consistent with the measurements from reprocessed R14 light curves (0.4σ lower, where σ is the statistical uncertainty from R14). The uncertainties on w are 15% higher and distance modulus uncertainties are 14% higher, likely due to the lower average S/N of photometric PS1 light curves. The median S/N at peak is 22 for all PS1 SNe, compared to a median S/N at peak of 38 for spectroscopically classified SNe.

6. Results from BEAMS Variants

The BEAMS method measures w with no significant bias due to CC SN contamination and a statistically insignificant bias in PS1 data. However, the reliability of these results could depend on the assumptions that we made when generating CC SN simulations and implementing BEAMS. We now expand our study of systematic uncertainties in simulations by applying alternative SN classification methods, including ones with less dependence on the accuracy of our CC SN simulations and adjusting the CC SN likelihood model.

6.1. Analysis Variants

In total, we test three additional methods of determining the prior probability $P(\text{Ia})$ (Equation (4))—the Nearest Neighbor (NN), Fitprob, and GalSNID classifiers—and two additional CC SN models. The two additional CC SN models include a two-Gaussian model and a single, asymmetric Gaussian model. NN and Fitprob are light-curve-based classification methods. NN uses SALT2 light-curve parameters to classify SNe based on whether they lie nearer to simulated SNe Ia or simulated CC SNe in X_1 , C , and redshift space while Fitprob uses the χ^2 and degrees of freedom of the SALT2 light-curve fit to measure a probability. GalSNID (Foley & Mandel 2013) uses the fact that, unlike CC SNe, many SNe Ia explode in galaxies with old stellar populations, and thus uses only host galaxy properties to derive an SN type probability. We expand the GalSNID method to use observables from host galaxy spectroscopy in addition to photometric observations. PSNID is the best method; NN yields a sample with 6.5% contamination at $P(\text{Ia}) > 0.5$ and 3.8% contamination at $P(\text{Ia}) > 0.9$, Fitprob yields a sample with 6.1% contamination at $P(\text{Ia}) > 0.5$ (4.1% at $P(\text{Ia}) > 0.9$), and GalSNID gives a sample with 9.3% contamination (7.2% at $P(\text{Ia}) > 0.9$; the total contamination in the sample is 9.7%). The details of these variants are given in Appendix B.

We note that the best approach would be a hybrid one that takes advantage of all classifiers. Though we keep these classifiers as separate here in order to explore the effect of different classification assumptions, Kessler & Scolnic (2017), for example, combine a Fitprob > 0.05 cut with the NN classifier. Combining GalSNID priors with a light-curve-based classifier is another promising option for future work.

We test each variant on 25 samples of 1000 simulated PS1 SNe. Though we discuss the ways in which distances and nuisance parameters are affected by these variants, we focus primarily on measurements of w . The rms of these variants gives an estimate of the systematic uncertainty on w , σ_w^{CC} , an error which could be reduced in the future by improved SN classification methods. It could also be reduced by testing our best single classifier on a robust set of CC+Ia SN simulations that include a larger set of CC SN templates and several methods of adjusting CC SN rates and LFs to match the data.

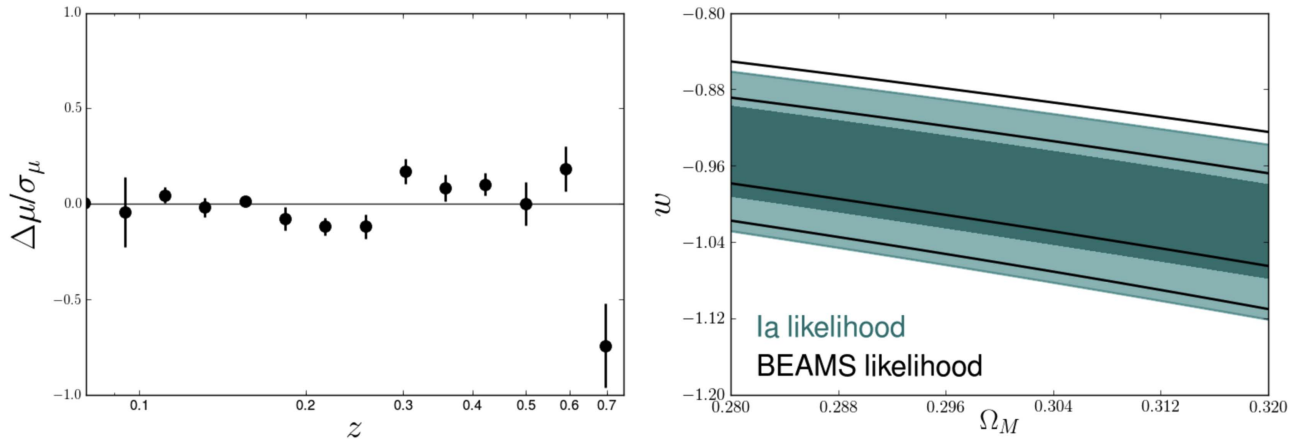


Figure 9. Left: average distance modulus bias due to CC SN contamination (Equation (7)) as a fraction of the statistical uncertainty. Error bars are the uncertainty on the median bias from 25 samples. The average absolute biases at $z > 0.2$ are ~ 3 mmag, with the point at $z \simeq 0.6$ having the largest bias of 6 mmag (with the exception of the final high-uncertainty point at $z \simeq 0.7$). There is a slight z -dependent slope, which could bias cosmological parameters, with at 2.3σ significance. Right: 1σ cosmological parameter likelihood contours from BEAMS compared to the true likelihood using a representative sample of 1000 PS1 SNe.

6.2. Systematic Uncertainty on w

We examine two situations in this section: one where α and β are measured by the BEAMS method, and one where α and β are fixed to the values measured from spectroscopic samples. In the case where α and β are measured by the BEAMS method, Figure 10 shows the bias on α and β from each classifier. β biases in particular can cause large distance biases at $z > 0.5$, as the average SN color at these redshifts is ~ -0.1 (for a bias in β of 0.2, $\Delta\beta \times \bar{C} = 20$ mmag).

If α and β are fixed, BEAMS requires very little information to give robust measurements of w . We test the effect of fixing α and β for all variants and also compare to the case where BEAMS has minimal prior information: we set $P(\text{Ia}) = 1/2$ for all photometric SNe while still fixing $P(\text{Ia}) = 1$ for low- z SNe. If α and β are fixed, the largest absolute bias on w is -0.018 (the Fitprob classifier) and the $P(\text{Ia}) = 1/2$ case gives a w bias of only -0.011 . The biases are approximately twice as high if we instead allow BEAMS to fit for α and β , and four times as high for the $P(\text{Ia}) = 1/2$ case, worse than all other methods (a w bias of -0.043).

Table 4 and Figure 11 show the median bias and increase in uncertainty on w due to each $P(\text{Ia})$ method and CC SN model. Figure 11 shows the bias before and after fixing α and β . We find that alternate CC SN models have only a small effect on the measurement of w . Our lowest w bias of -0.001 ± 0.003 comes from the skewed Gaussian CC SN model; however, the results from these three CC SN treatments are statistically consistent (with the exception of the two-Gaussian model with α and β fixed, which appears to have difficulty robustly measuring both CC SN Gaussians).

Using all variants, σ_w^{CC} has an average value of 0.014 ± 0.007 (30% of the statistical error) if α and β are fixed for the NN, GalSNID, and Fitprob classifiers (these classifiers give twice the bias on α and β as PSNID does). The uncertainty is due to the dispersion of the systematic uncertainty from sample to sample. BEAMS distances (Figure 12) and nuisance parameters (Figure 10) are consistent to within 1σ , regardless of the method.

We note that in some cases fixing α and β may subject the sample to additional systematic uncertainty. For example, α and β could be different in a photometric sample because the host galaxy spectroscopic follow-up selects bright hosts. Host

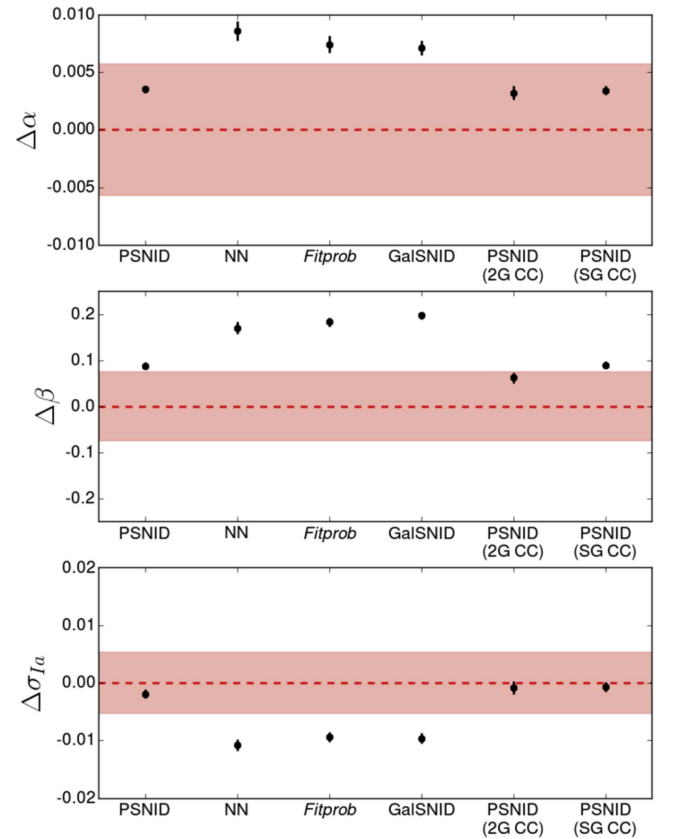


Figure 10. Bias in SALT2 α , β , and σ_{Ia} measured from 25 simulations of 1000 SNe each, with the shaded regions indicating typical uncertainties on each parameter from SN Ia-only samples. σ_{Ia} is too low by ~ 0.005 – 0.01 , while α and β are too high by ~ 0.005 – 0.01 and ~ 0.1 – 0.2 , respectively (~ 1 – 2σ). It is likely that reddened CC SNe are responsible for the higher color term (more consistent with Milky Way dust than the SN Ia color law). “2G CC” and “SG CC” refer to the two-Gaussian and skewed Gaussian CC SN parameterizations, respectively.

properties correlate with shape and color, which in turn can affect the measured α and β (Scolnic et al. 2014b). However, these biases are well-known and can in principle be simulated and corrected for (see Scolnic & Kessler 2016).

Current measurements of w (e.g., B14) have approximately equal statistical and systematic uncertainties. Therefore, a

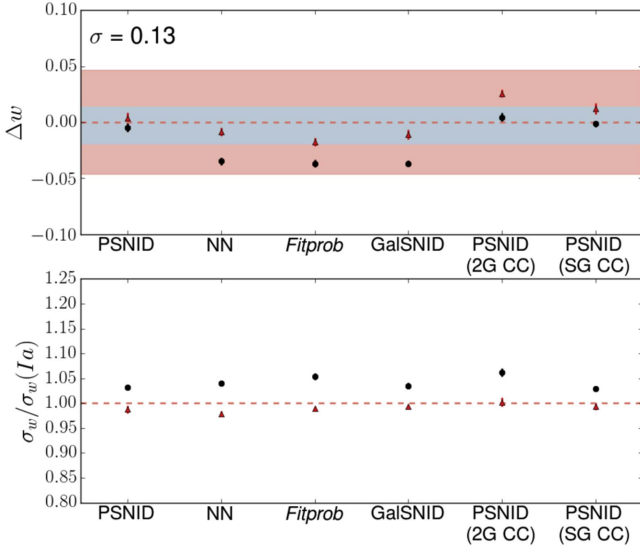


Figure 11. w bias (top) and increased uncertainties (bottom) due to different $P(Ia)$ priors and CC SN parameterizations in BEAMS (black points). We show the median from 25 samples of 1000 simulated SNe. Red points show the biases with α and β fixed. In the top panel, the statistical error on w from SNe Ia is shown in the red band and the dispersion of the values given in Table 4 in blue. Red points have lower uncertainties than the Ia-only uncertainties because fixing α and β neglects their uncertainties. “2G CC” and “SG CC” refer to the two-Gaussian and skewed Gaussian CC SN parameterizations, respectively.

Table 4
Cosmological Results from BEAMS Variants

Method	Δw^a	σ_{stat}^b	$\Delta w/\sigma_{\text{stat}}$	$\Delta\sigma_{\text{stat}}$
One Gaussian ^c	-0.005 ± 0.004	0.050	-0.1	0.002 (3%)
Two Gaussians ^c	0.004 ± 0.004	0.051	0.1	0.003 (6%)
Skewed Gaussian ^c	-0.001 ± 0.003	0.050	-0.0	0.001 (2%)
	Δw	$P(Ia)$ Method ^d σ_{stat}	$\Delta w/\sigma_{\text{stat}}$	$\Delta\sigma_w$
PSNID	-0.005 ± 0.004	0.050	-0.1	0.002 (3%)
NN ^e	-0.009 ± 0.004	0.047	-0.2	-0.001 (-2%)
Fitprob ^e	-0.018 ± 0.004	0.047	-0.4	-0.001 (-1%)
GalSNID ^e	-0.011 ± 0.004	0.048	-0.2	-0.000 (0%)

Notes. Bias of w in simulations from each CC SN model and prior probability method. We take the median of 25 samples of 1000 PS1 SNe. For each increase in uncertainty ($\Delta\sigma_w$), we show its percent increase in parentheses. Methods with the lowest bias are highlighted in bold.

^a The median bias on w and its uncertainty.

^b The statistical uncertainty on w from a single sample of 1000 PS1 SNe.

^c Using PSNID for the $P(Ia)$ prior probabilities.

^d Using a single-Gaussian CC SN model.

^e For these classifiers, we keep α and β fixed to their known values. $\Delta\sigma_w$ is negative in some cases, because fixing α and β neglects the contribution of nuisance parameter uncertainties to the uncertainty on w .

measurement of w biased by less than about half the statistical uncertainty (0.024 in this work), such as the value of $\sigma_w^{CC} = 0.014$ measured here, does not prohibit a robust measurement of w . Any bias larger than that—such as the alternative classifiers discussed in this section without α and β fixed—will dominate the systematic error budget and make it unlikely that photometric SN samples can be competitive with spectroscopically classified samples. For future surveys, such as DES and LSST, this bias may be approximately equal to the

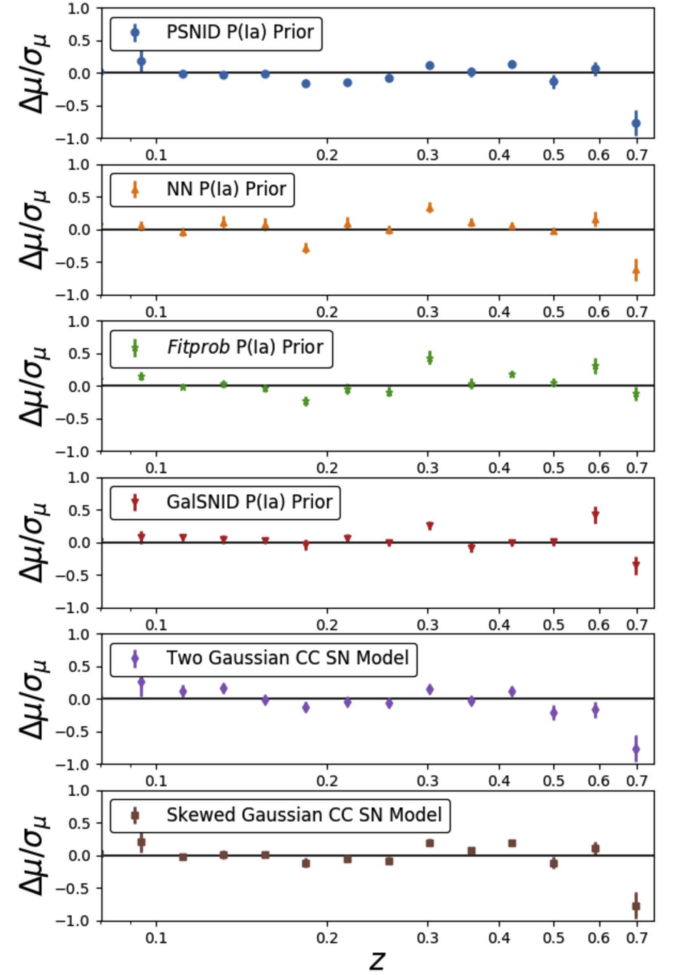


Figure 12. Distance bias due to CC SN contamination as a fraction of the distance uncertainty for each BEAMS variant. Small systematic discrepancies begin to appear at $z \geq 0.3$.

statistical error and must be reduced through improved classification methods or a better understanding of CC SNe to yield accurate results.

7. Discussion

The PS1 photometric SN sample is the largest SN Ia sample, but using it to optimally measure cosmological parameters—particularly if the nuisance parameters α and β are unknown or observationally biased—requires accurate SN type probabilities. These in turn rely on our understanding of the PS1 sample and the CC SNe in it. Evaluating how our incomplete knowledge of CC SNe could bias the results is difficult. In this section, we discuss how CC SN simulations could be improved in the future. We also present alternatives to our implementation of BEAMS and measure the degree to which different methods and priors affect the statistical uncertainty on w .

7.1. Generating Reliable CC SN Simulations

Evaluating the reliability of our method would be subject to fewer uncertainties if CC SN simulations were more robust. These simulations are currently subject to two primary limiting factors: the assumption that the CC SN LF is Gaussian with measured mean and rms from Li et al. (2011) and the limited CC SN template diversity.

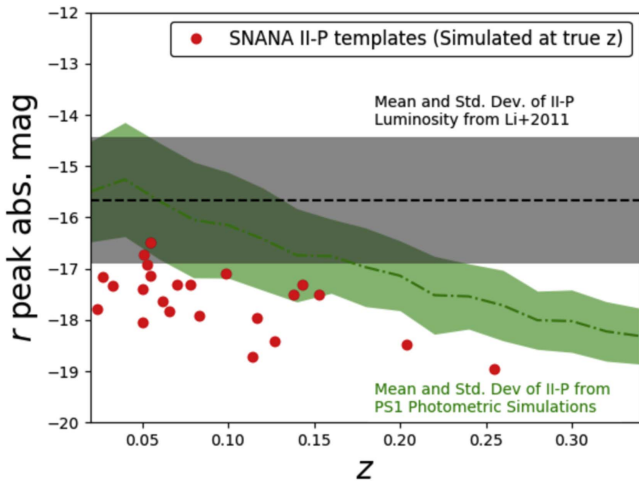


Figure 13. In this work, SNANA II-P templates from SDSS (red) are made fainter to match Li et al. (2011) LFs (gray) and then used to generate simulations of the PS1 survey (green). SNANA II-P templates are typically 2σ – 3σ brighter than the mean magnitude of the population from Li et al. (2011).

Figure 13 shows that the assumption of the shape of CC SN LFs could have a strong impact on the fraction of bright CC SNe. While the Malmquist bias for SNe Ia is ~ 0.1 mag at maximum, Type II SNe observed at the median PS1 survey redshift are up to 3 magnitudes—and 2–3 standard deviations—brighter than the peak of their LF. Determining the frequency of such bright CC SNe requires measuring the shapes of their LFs with better precision than what is currently available from volume-limited surveys such as Li et al. (2011). Due to low statistics, our current simulations treat the LFs of each SN subtype as Gaussian, a flawed assumption.

Generating more robust simulations also requires additional, diverse CC SN templates. Our simulations sample the luminosity, shape, and color distribution of most CC SN subtypes with just a few templates. In addition, the luminosity distribution of these templates is heavily biased; nearly all CC SNe currently used as templates are much brighter than the mean luminosity of their subtypes. Our method makes these bright templates fainter to match the Li et al. (2011) LFs, implicitly assuming that faint CC SNe have similar light curves to bright CC SNe. A better approach would be based on CC SN templates that sample the full range of luminosity space for CC SNe.

We note that additional high-S/N CC SN light curves and spectra exist, but require careful smoothing, interpolation, and spectral mangling to be a reliable addition to the SNANA template library. We have added SNe Ia-91bg and SNI Ib templates to SNANA (Appendix A.1.1), but assembling and mangling all available CC SN light curves and templates is beyond the scope of this work.

In the absence of additional templates and improved LF measurements, we can use GalSNID and Fitprob classifications to give measurements of w some degree of independence from these sources of uncertainty. Though these classifiers are sub-optimal compared to classifiers such as PSNID, they give a unique set of probabilities that do not rely on simulations for training (though Fitprob is implicitly dependent on the nature of CC SN light curves contaminating our sample). Fitprob and GalSNID *explicitly* depend on simulations only through their rates priors. Adjusting these priors by a factor of 2 biases w by $\sim 20\%$ of the statistical uncertainty or less.

7.2. Alternatives in Implementing BEAMS

In determining cosmological parameters with the BEAMS method, we made a set of choices with a modest number of free parameters that reproduced the full cosmological parameter likelihoods. We found that most choices, e.g., varying priors or adding additional CC SN bins, made little difference, provided that we had a large number of MCMC steps and few enough parameters.

Two additional choices can improve the systematic error due to CC SN contamination. First, though fixing α and β does not improve the accuracy of the BEAMS method when using PSNID priors, it does improve the accuracy when using the NN, Fitprob, and GalSNID methods with less accurate classifications. With α and β fixed, NN, Fitprob and GalSNID are competitive with the more sophisticated light-curve-based methods. If we choose to either keep α and β fixed when measuring w from these classifiers, we find that the σ_w^{CC} decreases by $\sim 30\%$ on average. In Pan-STARRS, spectroscopically confirmed SNe can measure these parameters with low uncertainty, and fixing them for our future cosmology analysis in some or all methods could be advantageous.

The second method of improving BEAMS is by cutting additional likely CC SNe from the sample. Following Kessler & Scolnic (2017), we tested a cut on the NN prior probability by requiring $0.5 < P_{NN}(\text{Ia}) < 1$. Our simulations show that this cut removes $\sim 33\%$ of contaminants but just 5% of SNe Ia. The rejected sample has $\sim 40\%$ CC SN contamination. We found that an NN probability cut yields no improvements to our results using the NN classifier. However, when this cut is added to our other classification methods, it reduces σ_w^{CC} by $\sim 30\%$ on average. We have not included this cut in our systematic error analysis (Section 6) as it makes our classification methods more correlated and adds an additional dependence on uncertain simulations to the measured systematic error. However, it is likely that this cut will increase the consistency of the full PS1 cosmological results. Kessler & Scolnic (2017) use a hybrid classification approach by requiring $\text{Fitprob} > 0.05$. In our simulations, this cut reduces the CC SN contamination by an additional 30% compared to using the NN classifier alone.

A third option for BEAMS is to estimate SN Ia distances with a stricter CC SN model. Kessler & Scolnic (2017) adopt an approach where BEAMS CC SN distributions are determined directly from simulations. For our PS1 analysis, we have adopted a more general approach to CC SNe at the cost of several additional parameters to marginalize over and a simpler form of the likelihood (Kessler & Scolnic 2017 also suggest free CC SN parameters as a possible improvement to their method). Tests show our parameterization is capable of marginalizing over the simulated CC SNe such that the Ia likelihood is recovered, and our method is slightly more general than a simulation-based method. A simulation-based mapping of CC SNe may be more robust, but validating it thoroughly is beyond the scope of this paper. In particular, the influence of inaccurate simulations on its recovered results must be explored fully.

7.3. Uncertainties in BEAMS Distances

By setting $P(\text{Ia}) = 1/2$ for all photometric SNe, the BEAMS method measures w with a bias of -0.01 , 0.2 times the statistical uncertainty on w . The statistical uncertainty on w from setting $P(\text{Ia}) = 1/2$, even with no prior information as to

which SNe are of Type Ia, is just 5% higher than using SNe Ia alone (comparing to SNe Ia alone in the case where α and β are fixed to known values). This is primarily due to two factors: the loose priors we employ and the fact that we include a sample of low- z spectroscopically confirmed SNe Ia for which $P(\text{Ia})$ is fixed to 1. These low- z SNe Ia help to set the SN Ia dispersion and the SN parameters α and β , which are fixed as a function of redshift.

If we remove the low- z sample, the distance and SN parameter biases increase. Distance uncertainties, which are higher by just $\sim 5\%$ when using the $P(\text{Ia}) = 1/2$ prior, increase by nearly 50%. Nevertheless, BEAMS does remarkably well at determining the Gaussian distributions of SNe Ia and CC SNe with relatively little information. This is helped by the fact that because SNe Ia have a factor of ~ 20 lower dispersion than CC SNe, a loose prior on BEAMS free parameters is sufficient to find the most probable Gaussian distributions.

If we use a more flexible CC SN model (a two-Gaussian or skewed Gaussian CC SN model), the requirements on our prior probabilities must become more stringent to yield precise distances. In the case of the two-Gaussian model, prior probabilities can no longer be renormalized or shifted (Equation (6))—these are parameters that can greatly improve the results for alternative prior probability methods. Second, our prior probabilities must be significantly more accurate to yield results with low uncertainties. With the two-Gaussian CC SN model, the uncertainty on w increases by 20% when using GalSNID priors and by 100% when setting $P(\text{Ia}) = 1/2$ for all photometric SNe. Using the skewed Gaussian model, the GalSNID and $P(\text{Ia}) = 1/2$ priors increase the uncertainties by 13% and 27%, respectively.

Fortunately, a single-Gaussian model for CC SNe appears to yield unbiased distances even though the simulated distribution is not perfectly Gaussian. In essence, BEAMS attempts only to determine the Gaussian distributions of two types of SNe and fortunately, those distributions are relatively well-separated in dispersion even if they are not always well-separated in distance.

8. Conclusions

We measured spectroscopic redshifts for 3147 SN host galaxies in Pan-STARRS, over 1000 of which are cosmologically useful, likely SNe Ia. When combined with the full PS1 spectroscopic sample (D. M. Scolnic et al. 2017, in preparation), we will have 1145 cosmologically useful SNe Ia from PS1.

We find that currently available CC SN templates and luminosity functions are biased or incomplete. Our results suggest there are too few bright CC SNe in our simulations.

We generate 25 simulations that closely resemble the PS1 sample. Each has 1000 photometric PS1 SNe and 250 low- z spectroscopically confirmed SNe Ia. These simulations show that our method can measure w with a bias due to CC SN contamination as low as -0.001 ± 0.003 . This equates to a systematic uncertainty on w of just 0.004, 8% of the statistical uncertainty, but this uncertainty could be affected by incomplete knowledge of the CC SN distribution. The SN Ia dispersion, σ_{Ia} , is biased by -0.005 ($\sim 0.5\sigma$), the SALT2 shape parameter α is biased by ~ 0.005 ($\sim 1\sigma$), and the color parameter β is biased by ~ 0.1 ($\sim 1.5\sigma$). The statistical uncertainties on w are nearly equivalent to those using only SNe Ia.

Using several variants of the method and a CMB-like prior on Ω_M , we estimate the systematic error introduced by CC SN contamination to be 0.014 ± 0.007 (29% of the statistical error). This systematic error would constitute only a 3% increase on the uncertainty on w in a JLA-like analysis with CMB priors ($\sigma_w = 0.057$ (stat+sys), and $\sqrt{0.057^2 + 0.014^2} = 0.059$). However, this systematic error assumes that α and β can be fixed to known values from a spectroscopic sample for the alternate classification methods. If α and β are fixed, our least accurate classifiers—including an uninformative prior probability $P(\text{Ia}) = 1/2$ for all simulated PS1 SNe—give a median bias on w between -0.01 and -0.02 . Systematic error could be reduced further by using a cut on prior probabilities from one variant to reduce CC SNe in the sample for the other variants. We caution that due to uncertainties in CC SN simulations and statistical fluctuations, the CC SN contamination systematic affecting our forthcoming cosmological results may be somewhat lower or higher than the one estimated in this work. However, that analysis will also include a subset of PS1 SNe with known (spectroscopic) classifications as part of the data, a scenario that will likely reduce the systematic uncertainty due to CC SN contamination.

Included in these variants are a total of four different classification methods to measure cosmology, including a host galaxy spectrum-based version of GalSNID (Foley & Mandel 2013) that we introduce in this work (see Appendix B.1.3). GalSNID is based only on SN Ia host galaxy observables and a rates prior. GalSNID provides a method of measuring w from photometric data that does not depend on SN light curves and training on simulated data. Machine learning techniques may be able to improve on the efficiency of this method in the future. We caution that even with these multiple variants, if CC SN simulations are inaccurate it could cause the systematic error to be underestimated in real data. Additional CC SN templates and a better measurement of the shape of CC SN LFs could help to ameliorate these concerns.

By drawing random samples from real PS1 data, we tested whether the BEAMS method can work on real data within the confidence intervals of Rest et al. (2014). We found that our measurements of w were fully consistent with those of Rest et al. (2014), as were the SN nuisance parameters α and β .

Though our results are robust, w is an extremely sensitive measurement and the burden of proof for BEAMS is high. Future validation tests could include SDSS and SNLS photometric data, as well as simulated tests with a variety of CC SN LFs. Additional light-curve classification methods could also help improve the reliability of the BEAMS method.

Future SN Ia samples from DES and LSST will be unable to rely solely on spectroscopic classification to measure cosmological parameters. With the light-curve classification and Bayesian methodologies presented here, we validate some of the techniques that will be used in future surveys, and anticipate that PS1 photometric SNe can provide a robust measurement of w using the largest SN Ia sample to date.

We would like to thank the anonymous referee for many helpful suggestions. This manuscript is based upon work supported by the National Aeronautics and Space Administration under Contract No. NNG16PJ34C issued through the *WFIRST* Science Investigation Teams Programme. R.J.F. and D.S. were supported in part by NASA grant 14-WPS14-0048. The UCSC group is supported in part by NSF grant

AST-1518052 and from fellowships from the Alfred P. Sloan Foundation and the David and Lucile Packard Foundation to R.J.F. This work was supported in part by the Kavli Institute for Cosmological Physics at the University of Chicago through grant NSF PHY-1125897 and an endowment from the Kavli Foundation and its founder Fred Kavli. D.S. gratefully acknowledges support from NASA grant 14-WPS14-0048. D.S. is supported by NASA through Hubble Fellowship grant *HST*-HF2-51383.001 awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS 5-26555.

Many of the observations reported here were obtained at the MMT Observatory, a joint facility of the Smithsonian Institution and the University of Arizona. This paper uses data products produced by the OIR Telescope Data Center, supported by the Smithsonian Astrophysical Observatory. Additional data are thanks to the Anglo-Australian Telescope, operated by the Australian Astronomical Observatory, through the National Optical Astronomy Observatory (NOAO PropID: 2014B-N0336; PI: D. Jones). We also use data from observations at Kitt Peak National Observatory, National Optical Astronomy Observatory, which is operated by the Association of Universities for Research in Astronomy (AURA) under a cooperative agreement with the National Science Foundation. Also based on observations obtained with the Apache Point Observatory 3.5 m telescope, which is owned and operated by the Astrophysical Research Consortium.

The computations in this paper used a combination of three computing clusters. BEAMS analysis was performed using the University of Chicago Research Computing Center and the Odyssey cluster at Harvard University. We are grateful for the support of the University of Chicago Research Computing Center for assistance with the calculations carried out in this work. The Odyssey cluster is supported by the FAS Division of Science, Research Computing Group at Harvard University. Supernova light-curve reprocessing would not have been possible without the Data-Scope project at the Institute for Data Intensive Engineering and Science at Johns Hopkins University.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Center for High-Performance Computing at the University of Utah. The SDSS Web site is <http://www.sdss.org>.

SDSS-IV is managed by the Astrophysical Research Consortium for the Participating Institutions of the SDSS Collaboration including the Brazilian Participation Group, the Carnegie Institution for Science, Carnegie Mellon University, the Chilean Participation Group, the French Participation Group, Harvard-Smithsonian Center for Astrophysics, Instituto de Astrofísica de Canarias, The Johns Hopkins University, Kavli Institute for the Physics and Mathematics of the Universe (IPMU)/University of Tokyo, Lawrence Berkeley National Laboratory, Leibniz Institut für Astrophysik Potsdam (AIP), Max-Planck-Institut für Astronomie (MPIA Heidelberg), Max-Planck-Institut für Astrophysik (MPA Garching), Max-Planck-Institut für Extraterrestrische Physik (MPE), National Astronomical Observatory of China, New Mexico State University, New York University, University of Notre Dame, Observatório

Nacional/MCTI, The Ohio State University, Pennsylvania State University, Shanghai Astronomical Observatory, United Kingdom Participation Group, Universidad Nacional Autónoma de México, University of Arizona, University of Colorado Boulder, University of Oxford, University of Portsmouth, University of Utah, University of Virginia, University of Washington, University of Wisconsin, Vanderbilt University, and Yale University.

This research makes use of the VIPERS-MLS database, operated at CeSAM/LAM, Marseille, France. This work is based in part on observations obtained with WIRCam, a joint project of CFHT, Taiwan, Korea, Canada, and France. The CFHT is operated by the National Research Council (NRC) of Canada, the Institut National des Sciences de l'Univers of the Centre National de la Recherche Scientifique (CNRS) of France, and the University of Hawaii. This work is based in part on observations made with the *Galaxy Evolution Explorer* (GALEX). GALEX is a NASA Small Explorer, whose mission was developed in cooperation with the Centre National d'Etudes Spatiales (CNES) of France and the Korean Ministry of Science and Technology. GALEX is operated for NASA by the California Institute of Technology under NASA contract NAS5-98034. This work is based in part on data products produced at TERAPIX available at the Canadian Astronomy Data Centre as part of the Canada–France–Hawaii Telescope Legacy Survey, a collaborative project of NRC and CNRS. The TERAPIX team has performed the reduction of all the WIRCAM images and the preparation of the catalogs matched with the T0007 CFHTLS data release.

Funding for the DEEP2 Galaxy Redshift Survey has been provided by NSF grants AST-95-09298, AST-0071048, AST-0507428, and AST-0507483 as well as NASA LTSA grant NNG04GC89G. This research uses data from the VIMOS VLT Deep Survey, obtained from the VVDS database operated by Cesam, Laboratoire d'Astrophysique de Marseille, France. zCosmos data are based on observations made with ESO Telescopes at the La Silla or Paranal Observatories under program ID 175.A-0839.

Appendix A The Dearth of Simulated CC SNe

A.1. Core-collapse SNe

There are a few potential explanations for the difference in Hubble residuals ($0.5 < \mu - \mu_{\Lambda\text{CDM}} < 1.5$) between simulations and data. In this appendix, we attempt to identify the cause of the discrepancy.

First, a large percentage ($\gtrsim 20\%$) of inaccurate SNIa redshifts could explain the data. However, in addition to disagreeing with our measurements, this would give too many simulated SNe with very bright and very faint Hubble residuals. Requiring a high TDR minimum and a small separation between the SN location and host galaxy center in our data does not resolve the conflict.

A second option is that the relative rates or magnitude distributions from Li et al. (2011) are erroneous or are biased by the targeted nature of the survey (LOSS searched for SNe in a set of pre-selected bright galaxies). These rates also do not take into account that the relative fractions of different CC SN subtypes could change with redshift. Modest adjustments, such as “tweaking” the mean magnitudes or dispersions of CC SNe by $\lesssim 0.5$ mag, cannot explain the discrepancy. Simulating

CC SNe using LFs from Richardson et al. (2014), which are typically ~ 0.3 – 1.0 mag brighter than those of Li et al. (2011), produces far too many bright CC SNe compared to our data. The effect of weak lensing on the data is expected to be an order of magnitude less than the size of the offset we see here (Smith et al. 2014). It is also unlikely that strongly lensed SNe contribute significantly to the discrepancy (Oguri & Marshall 2010).

By reclassifying LOSS SNe, Shivvers et al. (2017) recently found that SN Ib relative rates were more than double the fraction found by Li et al. (2011). This change could reduce the Hubble residual discrepancy by half or more. However, Shivvers et al. (2017) determined these rates by reclassifying a number of LOSS SNe Ic as SNe Ib, which in turn means that the SN Ib LF should be made fainter. Making the SN Ib LF fainter will increase the discrepancy in Hubble residuals. We continue to use Li et al. (2011) in this work, as we can be sure that the LFs and relative rates are self-consistent.

Finally, we consider that our results could be biased if SNANA templates have lower average reddening than PS1 data. There are likely substantial differences between the reddening distribution of the templates and the data. However, we find that adding additional reddening to our simulations tends to make the magnitude distribution of CC SNe broader (we approximately adjust the Li et al. 2011 LFs for dust following Rodney et al. 2014). This *increases* the discrepancy between simulations and data. Correcting for the unknown intrinsic reddening of these templates is an important future objective that can allow SNANA simulations to be more realistic. See Section 7.1 for further discussion of biases in our simulations and templates.

A.1.1. Adding New Supernova Templates to SNANA

Several CC SN or peculiar Ia subtypes are missing from the SNANA simulation library but could be present in the PS1 data. Missing SN types include superluminous SNe, SNe Iib, SNe Ibc-pec, and peculiar, faint SNe Ia such as 1991bg-like SNe Ia (Ia-91bg) and SNe Iax (Foley et al. 2013). Superluminous SNe are unlikely to help resolve the discrepancy, as they are brighter than SNe Ia and occur preferentially in faint hosts for which redshifts are difficult to measure (Lunnan et al. 2015). SNe Ibc-pec have similar LFs to SNe II-P but are much less common, so it is unlikely that many would fall on the Hubble diagram so near the SN Ia distribution. SNe Iax are red, fast-declining SNe that may be relatively common but have faint (albeit uncertain) LFs more similar to SNe II-P and Ibc-pec. These also tend to be poorly fit by SALT2, and would frequently fail our cuts.

SNe Iib and SNe Ia-91bg both have LFs only ~ 1 mag fainter than SNe Ia, though they are relatively uncommon and would need a high fraction to pass SALT2 light-curve cuts to be major contributors to our Hubble diagram. We investigated their impact by adding Ia-91bg and Iib templates to SNANA.

To simulate CC SNe over a wide range of redshifts and passbands, SNANA templates require relatively high-S/N, high-cadence spectral and photometric sampling, which exists for a paucity of CC SNe. Simulating SN light curves at high redshift often necessitates near-ultraviolet data as well. To create a template, an interpolated, flux-calibrated spectral time series is “mangled” to match the observed photometry by using wavelength-dependent splines with knots at the effective wavelengths of the photometric filters. Least-squares fitting

determines the best-fit spline that scales the spectrum to match the photometry. Hsiao et al. (2007) describes the “mangling” procedure in detail.

To improve the SNANA CC SN simulation, we add four SN Iib templates—SNe 1993J, 2008ax, 2008bo, and 2011dh—using spectra and light curves consolidated by the Open Supernova Catalog (Guillochon et al. 2017).²¹ Each of these templates have well-sampled spectra and optical light curves. We also add Ia-91bg templates using the SN 1991bg spectrum from Nugent et al. (2002)²², warped to match SNe Ia-91bg with well-sampled light curves before and after maximum (SNe 1991bg, 1998de, 1999by, 2005bl.²³). Using multiple SN templates helps us obtain better sampling of the shape—luminosity relation for SNe 91bg (steeper than the relation for normal SNe Ia; Taubenberger et al. 2008).

Figure 14 shows the interpolated light curves, mangled spectra, and Hubble residual histograms for SNe Iib and Ia-91bg. For Ia-91bg, we assume their rates have the same redshift dependence as SNe Ia. SNe Ia-91bg have magnitude distributions that could explain the data, but their rates are inconsistent with the data. SNe Iib are far too rare, as nearly all simulated SNe Iib have measured colors that are too red to be SNe Ia. Though we find that Ia-91bg and Iib SNe are not frequent enough to resolve the difference between PS1 data and simulations, we incorporate these subtypes in our simulations hereafter.

A.1.2. Measuring CC SN LFs with PSNID

There is an additional procedure by which PS1 data can inform CC SN LFs: we use the PSNID light-curve classifier (Sako et al. 2011, 2014) to separate the likely contributions of SNe Ia, Ib/c, and II. The SNANA implementation of PSNID compares the SALT2 SN Ia model and SNANA’s CC SN templates to the observed data. PSNID determines the fit χ^2 - and prior-based probability that a given SN is Type Ia, Type Ib/c, or Type II. Though the set of templates we use for PSNID is the same set we use to generate CC SNe in our simulations, broad priors allow these templates to be shifted in magnitude and extinction to fit our data.

We compare PSNID’s classifications of PS1 data and simulations by examining the distribution of $m_B - \mu_{\Lambda\text{CDM}}$, a proxy for absolute magnitude at peak (Figure 15). We find that likely SNe Ib/c are much brighter and have lower dispersion than the simulations. To bring our simulations into agreement with the data, we adjusted the simulated SN Ib/c and II distributions such that the mean and standard deviations of the *simulated* SNe that PSNID classified as Type Ib/c and II matched the mean and standard deviations of the *real* PS1 SNe that PSNID classified as Type Ib/c and II. This requires

²¹ References for the spectra and photometry are listed here. SN 1993J: Richmond et al. (1996), Metlova et al. (1995), Barbon et al. (1995), Jerkstrand et al. (2015), and Modjaz et al. (2014). SN 2008ax: Modjaz et al. (2014), Brown et al. (2014), Taubenberger et al. (2011), Tsvetkov et al. (2009), and Pastorello et al. (2008). SN 2008bo: Modjaz et al. (2014), Brown et al. (2014), and Bianco et al. (2014). SN 2011dh: Ergon et al. (2015, 2014), Shivvers et al. (2013), and Arcavi et al. (2011). Secondary sources: Yaron & Gal-Yam (2012), Richardson et al. (2001), Silverman et al. (2012), and the Sternberg Astronomical Institute Supernova Light Curve Catalogue.

²² https://c3.lbl.gov/nugent/nugent_templates.html

²³ References for the photometry are listed here. SN 1998de: Silverman et al. (2012), Ganeshalingam et al. (2010), and Modjaz et al. (2001). SN 1999by: Silverman et al. (2012), Ganeshalingam et al. (2010), and Garnavich et al. (2004). SN 2005bl: Contreras et al. (2010). Secondary sources: the Sternberg Astronomical Institute Supernova Light Curve Catalogue.

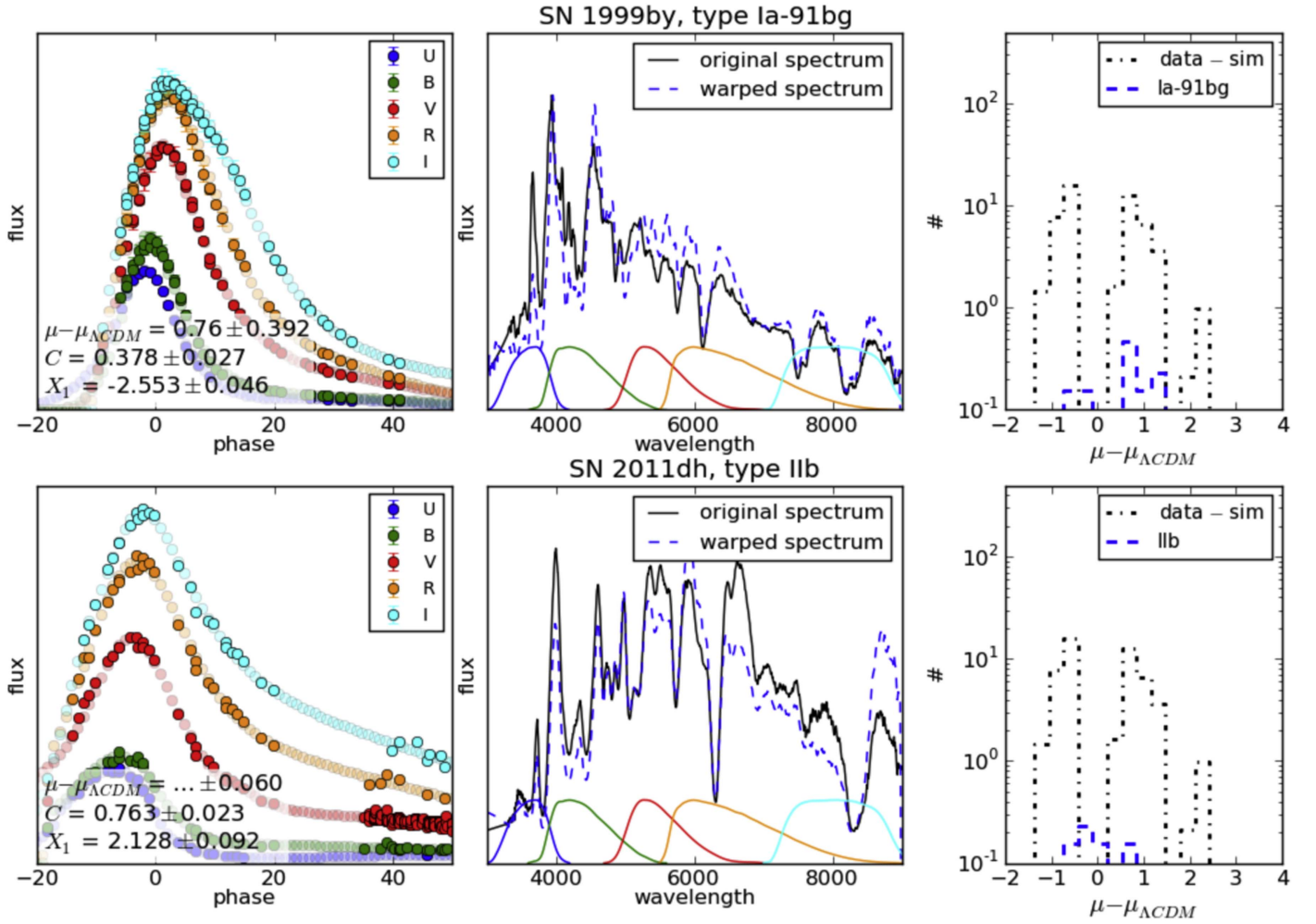


Figure 14. New templates for SNe Ia-91bg (top panels) and IIb (bottom panels) were added to SNANA by mangling a template spectrum to match light-curve data. From left to right, we show the interpolated SN light curves (light shading indicates interpolated points), the warped template spectra at peak brightness, and the Hubble residuals of *all* templates of the new subtype. We compare the Hubble residuals of the data and our simulations; the new templates cannot explain the discrepancy we observe. Because SN 2011dh has $z < 0.01$, its distance modulus residual is not shown in the left panel.

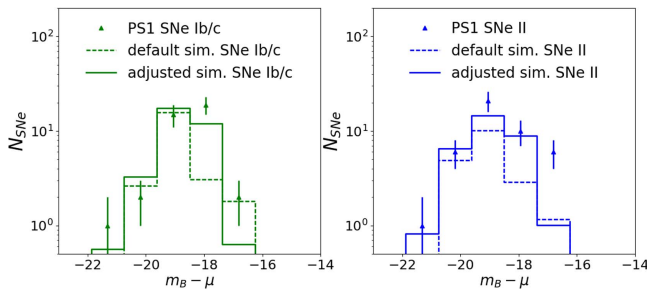


Figure 15. Empirical adjustments to SNANA simulations motivated by PSNID classifications, shown using histograms of SN absolute magnitude (SALT2 $m_B - \mu_{\text{ACDM}}$). PSNID-classified PS1 SNe and PSNID-classified simulations suggest that SNe Ib/c, after shape and color cuts, are brighter than expected. Our adjusted simulations (solid lines) match the data after we reduce the simulated dispersions and brighten LFs by ~ 1 mag.

reducing the dispersion of CC SN templates by 55% for SNe Ib/c. It also requires brightening the simulated LFs by 1.2 mag for SNe Ib/c and 1.1 mag for SNe II. We made shape and color cuts (Section 2.2) in this analysis but neglected σ_{X_1} and σ_{peakMJD} cuts to increase our SN statistics.

Figure 15 shows the distributions of PSNID-classified PS1 SNe ($P(\text{SN Type}) > 95\%$) compared to our simulations before

and after absolute magnitude and dispersion adjustments. We apply shape and color cuts but neglect additional cuts to increase our CC SN sample size.

After these adjustments, simulated CC+Ia SNe are consistent with our data. Figure 5 shows Hubble residual histograms before and after our PSNID-based adjustments. After correction, CC SNe are 8.9% of our final sample and SNe Ia-91bg comprise 0.2%. Additional CC SNe can explain the red tail of the SALT2 C distribution in Figure 4(C) (Figure 16). No CC SN rate adjustments were made. Although the simulated absolute magnitudes have been brightened by ~ 1 mag, CC SN in the adjusted simulations are only ~ 0.5 mag brighter than the original simulation on average. This is because as we brighten the CC SN distribution, the number of detectable faint SNe—which are nearer to the peak of the LF, and thus occur more frequently—increases, reducing the mean absolute magnitude. Note that the $\sim 2\sigma$ – 3σ discrepancy on the left (bright) side of the Hubble diagram can be reduced by simulating a nominal host mass correction, which tends to very slightly broaden the simulated distribution of SNe Ia.

Our adjusted simulation matches the Hubble residuals of the PS1 data. It also resolves the discrepancies in the PS1 C distribution (Figure 16). Hereafter, we refer to the adjusted

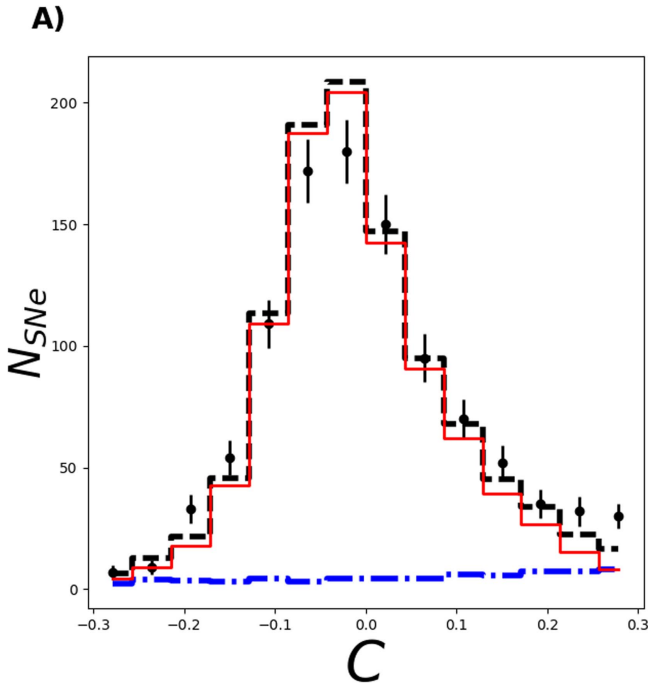


Figure 16. Simulated SALT2 C (right), compared to data in the J17 (adjusted) simulations. Compared to the original simulations (Figures 4(D) and (E)), the red end of the C distribution is more consistent with our data in the J17 simulations.

simulation, which adds new CC SN templates and uses PSNID to infer the true SN Ib/c distribution, as the J17 simulation.²⁴

Appendix B Expanding the BEAMS Method

We discuss the methodology behind alternative BEAMS variants in this section. The results from these variants are given in Section 6.

B.1. Additional $P(\text{Ia})$ Priors

In addition to the PSNID prior probabilities in our baseline method, we use three additional methods of estimating $P(\text{Ia})$: Fitprob, NN, and GalSNID. The effectiveness of each method is illustrated in Figure 17. The NN, Fitprob, and PSNID classifiers all determine probabilities by fitting to the photometric SN light curve. Fitprob relies on only the SALT2 model for fitting, while PSNID and NN depend on CC SN simulations for templates and training, respectively. GalSNID uses host galaxy information and depends on SNANA simulations only through the SN rates prior.

B.1.1. Nearest Neighbor

The NN classifier (Sako et al. 2014) uses a set of observables to define how close a given SN is to the CC SN and SN Ia populations. In our implementation, we use the SALT2 color (C), stretch (X_1), and redshift (z). The equation

$$d_i^2 = \frac{(z - z_i)^2}{\Delta z_{\max}^2} + \frac{(C - C_i)^2}{\Delta C_{\max}^2} + \frac{(X_1 - X_{1,i})^2}{\Delta X_{1,\max}^2} \quad (9)$$

²⁴ Templates and simulation input files for this simulation have been added to the SNANA library.

defines a list of NN distances between the i th SN and simulated training data. For the i th SN, neighbors are defined as all simulated events with $d_i < 1$. NN training finds the parameters ΔC_{\max} , $\Delta X_{1,\max}$, and Δz_{\max} that optimize the classification metric (efficiency \times purity) of simulated training data. NN is an efficient and accurate classifier in PS1 simulations: the set of SNe with $P_{\text{NN}}(\text{Ia}) > 0.9$ has 3.8% contamination compared to 9.7% contamination for the full sample (including CC SNe and SNe Ia with incorrect redshifts). This set includes 74% of all SNe Ia. See Kessler & Scolnic (2017) for details on the NN classification method.

B.1.2. Fitprob

The Fitprob method estimates $P(\text{Ia})$ from the χ^2 and number of degrees of freedom of the SALT2 light-curve fit (the SALT2 fit probability). Because the SALT2 fit χ^2 has no knowledge about the relative frequency of different SN types, we multiplied $P_{\text{fp}}(\text{Ia})$, the Fitprob probability, by a redshift-dependent SN rates prior, $P(\text{Ia}|z)$. $P(\text{Ia}|z)$ is the number of SNe Ia divided by the total number of SNe at a given redshift (after sample cuts; measured using the J17 simulations):

$$\tilde{P}_{\text{fp}}(\text{Ia}) = \frac{P(\text{Ia} | z) P_{\text{fp}}(\text{Ia})}{P(\text{CC} | z)(1 - P_{\text{fp}}(\text{Ia})) + P(\text{Ia} | z) P_{\text{fp}}(\text{Ia})} \quad (10)$$

$$P(\text{CC}|z) = 1 - P(\text{Ia}|z).$$

Compared to the PSNID (baseline) classifier, Fitprob has twice the fraction of contaminants at $P(\text{Ia}) > 0.5$. The fraction of CC SNe with high $P(\text{Ia})$ is also higher by a factor of ~ 2 .

B.1.3. GalSNID

SNe Ia have much longer average delay times between progenitor formation and explosion than CC SNe. Because of this, SNe Ia are the only SN type found in early-type hosts. This allows methods such as GalSNID (Foley & Mandel 2013) to classify SNe with host galaxy information. The GalSNID method in Foley & Mandel (2013) is based on photometric information and is highly dependent on host morphology. Because measuring galaxy morphologies at typical PS1 redshifts requires $\sim 0''.1$ image resolution, we modified the method by adding spectral observables. Though GalSNID is a very inefficient classifier, it measures SN Ia probabilities in a way that is only minimally subject to light-curve and LF uncertainties.

To train GalSNID, we used 602 host galaxy spectra from the Lick Observatory Supernova Search (LOSS; Leaman et al. 2011) and 354 host galaxy spectra of PS1 spectroscopically confirmed SNe. The equivalent widths of spectral emission lines, and $H\alpha$ in particular, correlate with SN type. Another useful diagnostic is the template that cross-correlates best with the observed host spectrum. Finally, we include host galaxy R (labeled effective offset in Foley & Mandel 2013), $B - K$, colors and absolute K magnitudes from Foley & Mandel (2013).

We trained GalSNID on spectral information using LOSS host galaxy spectra and spectroscopically confirmed PS1 SNe for which we have host galaxy spectra. Relative to the PS1 spectroscopic sample, LOSS has a greater number of total SNe, and a greater diversity and number of CC SNe on which to train the data.

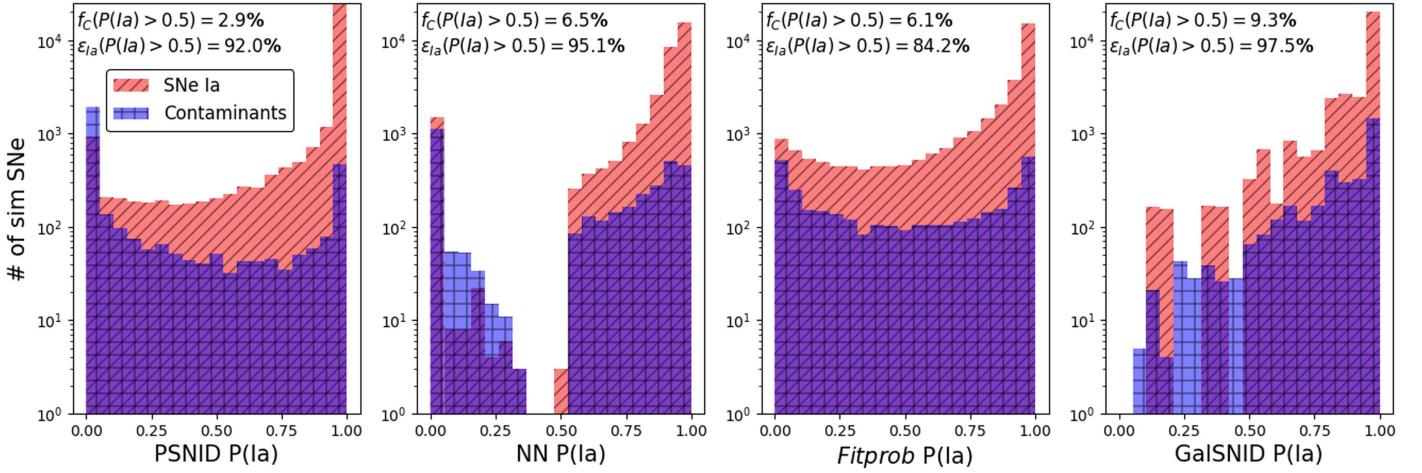


Figure 17. Simulated prior probabilities from the four classification methods discussed in this work for SNe Ia (red) and contaminants (blue; includes CC SNe and SNe Ia with incorrect redshifts). For each method, we show the percentage of contaminants f_C and the fraction of SNe Ia included, ϵ_{Ia} , in a $P(Ia) > 0.5$ sample.

Spectra for $\sim 1/3$ of the LOSS sample are available from SDSS/BOSS (Alam et al. 2015; 297 spectra), and we found an additional $\sim 1/3$ (305 spectra) by querying the NASA/IPAC Extragalactic Database. In total, 67% of the 905 SNe discovered by LOSS have host galaxy spectra. In general, the S/N of these data are high (much higher on average than our redshift survey data).

PS1 spectroscopically classified 520 SNe, of which ~ 150 are CC SNe and the rest are SNe Ia. Of the CC SNe, ~ 30 are SNe II_n (M. R. Drout et al. 2017, in preparation), 76 are II-P or II-L (Sanders et al. 2015), and ~ 20 –30 are SNe Ib or Ic. We obtained host galaxy spectra for 354 of these SNe.

We searched for a number of prominent, observational galaxy diagnostics that correlate with the age of the host, and found that the equivalent widths of bright emission lines such as O II, O III, H α , and H β are measurable in many of our spectra. We required continuum S/N > 5 near a given line measurement for an observable to be used in training or classification. As a way to incorporate additional information in a single diagnostic, we included the best-matched spectral template based on cross-correlation as an observable.

Although these diagnostics are correlated, in this work we follow Foley & Mandel (2013) in treating them as independent. Final probabilities for a given SN can therefore be computed by multiplying the probability of a Ia given each observable (Foley & Mandel 2013):

$$P(Ia|D) = k^{-1} P(Ia|z) \prod_{i=1}^N P(D_i|Ia), \quad (11)$$

where N is the number of observables and $P(D_i|Ia)$ is the probability of an observable given that the SN is Type Ia (Table 5). $P(D_i|Ia)$ is easy to compute; it is the fraction of SN Ia host galaxies that have observable D_i . $P(Ia|z)$ is a rates prior informed by our SNANA simulations. k is a normalization factor that requires $P(Ia|D) + P(CC|D) = 1$. See Foley & Mandel (2013) for additional details on the methodology. In the future, machine learning techniques may be able to improve our results by relaxing the assumption that observables are uncorrelated.

The probabilities from our LOSS+PS1 training sample are provided in Table 5. We also include the effective offset, $B - K$ colors, and K absolute magnitudes using probabilities

Table 5
Probability of Host Properties Given Type

Bin	$P(D_i Ia)$	$P(D_i Ibc)$	$P(D_i II)$
Cross-correlation Template			
Absorption	0.502 $^{+0.054}_{-0.048}$	0.256 $^{+0.069}_{-0.055}$	0.286 $^{+0.036}_{-0.033}$
Ellipt+A stars	0.431 $^{+0.050}_{-0.045}$	0.598 $^{+0.097}_{-0.085}$	0.609 $^{+0.051}_{-0.048}$
Late-type	0.029 $^{+0.017}_{-0.012}$	0.037 $^{+0.035}_{-0.020}$	0.030 $^{+0.015}_{-0.010}$
Emission	0.029 $^{+0.017}_{-0.012}$	0.098 $^{+0.047}_{-0.034}$	0.071 $^{+0.021}_{-0.016}$
H α Equivalent Width			
< -5.0	0.005 $^{+0.011}_{-0.004}$	0.000 $^{+0.033}_{-0.000}$	0.000 $^{+0.010}_{-0.000}$
$-5.0-0.0$	0.323 $^{+0.045}_{-0.039}$	0.054 $^{+0.051}_{-0.029}$	0.116 $^{+0.031}_{-0.024}$
$0.0-5.0$	0.219 $^{+0.038}_{-0.033}$	0.250 $^{+0.086}_{-0.066}$	0.217 $^{+0.039}_{-0.034}$
$5.0-10.0$	0.095 $^{+0.026}_{-0.022}$	0.125 $^{+0.067}_{-0.046}$	0.143 $^{+0.033}_{-0.027}$
> 10.0	0.358 $^{+0.047}_{-0.042}$	0.571 $^{+0.120}_{-0.100}$	0.524 $^{+0.058}_{-0.052}$
H β Equivalent Width			
< -5.0	0.000 $^{+0.007}_{-0.000}$	0.000 $^{+0.026}_{-0.000}$	0.000 $^{+0.007}_{-0.000}$
$-5.0-0.0$	0.504 $^{+0.046}_{-0.043}$	0.338 $^{+0.084}_{-0.068}$	0.333 $^{+0.040}_{-0.035}$
$0.0-5.0$	0.399 $^{+0.041}_{-0.038}$	0.451 $^{+0.094}_{-0.079}$	0.441 $^{+0.044}_{-0.041}$
$5.0-10.0$	0.069 $^{+0.019}_{-0.016}$	0.070 $^{+0.048}_{-0.030}$	0.149 $^{+0.028}_{-0.023}$
> 10.0	0.029 $^{+0.014}_{-0.010}$	0.141 $^{+0.060}_{-0.044}$	0.077 $^{+0.021}_{-0.017}$
O II Equivalent Width			
< -5.0	0.000 $^{+0.027}_{-0.000}$	0.000 $^{+0.183}_{-0.000}$	0.000 $^{+0.056}_{-0.000}$
$-5.0-0.0$	0.103 $^{+0.055}_{-0.038}$	0.000 $^{+0.183}_{-0.000}$	0.152 $^{+0.101}_{-0.066}$
$0.0-5.0$	0.676 $^{+0.115}_{-0.098}$	0.400 $^{+0.315}_{-0.191}$	0.545 $^{+0.161}_{-0.126}$
$5.0-10.0$	0.132 $^{+0.060}_{-0.043}$	0.300 $^{+0.290}_{-0.163}$	0.182 $^{+0.108}_{-0.072}$
> 10.0	0.074 $^{+0.049}_{-0.032}$	0.300 $^{+0.290}_{-0.163}$	0.121 $^{+0.096}_{-0.058}$
O III Equivalent Width			
< -5.0	0.000 $^{+0.007}_{-0.000}$	0.000 $^{+0.027}_{-0.000}$	0.000 $^{+0.007}_{-0.000}$
$-5.0-0.0$	0.215 $^{+0.032}_{-0.028}$	0.101 $^{+0.055}_{-0.037}$	0.079 $^{+0.021}_{-0.018}$
$0.0-5.0$	0.674 $^{+0.054}_{-0.050}$	0.739 $^{+0.118}_{-0.103}$	0.728 $^{+0.058}_{-0.053}$
$5.0-10.0$	0.067 $^{+0.019}_{-0.016}$	0.058 $^{+0.046}_{-0.028}$	0.059 $^{+0.019}_{-0.015}$
> 10.0	0.041 $^{+0.016}_{-0.012}$	0.101 $^{+0.055}_{-0.037}$	0.134 $^{+0.027}_{-0.023}$

measured from Foley & Mandel (2013) and SED fits using PS1 host galaxy photometry. Note that because H α and H β are almost perfectly correlated (the correlation coefficient is 0.94),

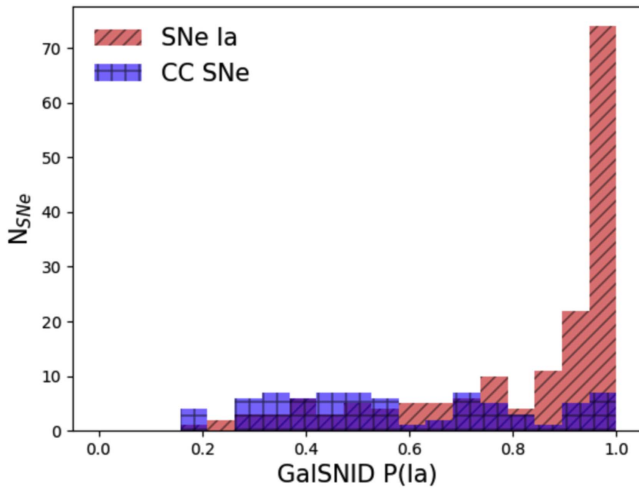


Figure 18. GalSNID classifications of spectroscopically classified CC SNe and SNe Ia in Pan-STARRS, neglecting rates priors.

we do not use $H\beta$ as an observable when $H\alpha$ is present in optical spectra ($z \lesssim 0.35$). Figure 17 shows the GalSNID probabilities of SNe Ia and CC SNe in PS1 and our simulations (we redshift and add noise to LOSS spectra to determine simulated GalSNID probabilities). Figure 18 shows GalSNID probabilities for real spectroscopically classified PS1 SNe.

To create GalSNID probabilities for the simulated sample, we artificially redshifted LOSS host galaxy spectra, added noise to make them consistent with the S/N of PS1 host spectra, and used GalSNID to measure the probability that each host observed an SN Ia. We took the distributions of GalSNID probabilities for the redshifted, noisy spectra corresponding to LOSS SNe II, Ib/c, and Ia hosts in each simulated redshift bin and assigned the probabilities drawn from those distributions to simulated SNe II, Ib/c, and Ia. This gave our simulated SNe the same probability distributions as the redshifted LOSS data. Figure 17 shows that GalSNID is a relatively imprecise classifier, but it provides constraints that are independent of SN light curves and their associated uncertainties. We have not taken into account the redshift evolution of SN host galaxies in this work.

On PS1 data, GalSNID is by far the least efficient classifier. Because classifications are highly influenced by the rates prior, GalSNID considers just 5% of contaminants to be likely CC SNe. If we set a higher threshold of $P(\text{Ia}) > 0.9$, GalSNID removes $\sim 25\%$ of CC SNe and keeps $\sim 70\%$ of SNe Ia. GalSNID is also most effective at $z \lesssim 0.35$, where $H\alpha$ is present in our optical spectra (the best indicator of SN type in our spectra). Unfortunately, the largest SN Ia distance biases are at $z > 0.4$, where the CC SN distribution becomes blended with the SN Ia distribution.

GalSNID would also be useful as an additional prior on SN type in conjunction with other methods. However, due to the uncertainty in CC SN models and LFs, in the present analysis we consider it most powerful as a stand-alone tool that can measure SN Ia probabilities without using light-curve data.

B.2. Varying the CC SN Model

PS1 and other spectroscopic data show that SNe Ia are well-represented by a Gaussian Hubble residual model, but CC SNe are not. We investigated replacing the CC SN likelihood in Equation (4) with two likelihoods that are more consistent with

our CC SN simulations. We tested a two-Gaussian model with 10 additional free parameters for CC SNe (the means and standard deviations of the second Gaussian at five redshift control points). We also tested a single, asymmetric Gaussian model with five additional free parameters (skewness at each CC SN control point).

If we allow BEAMS to shift and/or rescale the prior probabilities that an SN is of Type Ia (Equation (6)), BEAMS can give unphysical results. The alternative CC SN models are significantly more flexible and that flexibility must be constrained by accurate, fixed prior probabilities such as those from NN (see Section 7.3). We fix the parameters that allow BEAMS to adjust the priors ($A = 1$ and $S = 0$ in Equation (6)) or else the uncertainties on SN Ia distances will inflate to > 0.1 mag for even our best-measured redshift control points.

References

- Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, *ApJS*, **219**, 12
- Arcavi, I., Gal-Yam, A., Yaron, O., et al. 2011, *ApJL*, **742**, L18
- Barbon, R., Benetti, S., Cappellaro, E., et al. 1995, *A&AS*, **110**, 513
- Bernstein, J. P., Kessler, R., Kuhlmann, S., et al. 2012, *ApJ*, **753**, 152
- Betoule, M., Kessler, R., Guy, J., et al. 2014, *A&A*, **568**, A22
- Bianco, F. B., Modjaz, M., Hicken, M., et al. 2014, *ApJS*, **213**, 19
- Blake, C., Brough, S., Couch, W., et al. 2008, *A&G*, **49**, 5.19
- Brown, P. J., Breeveld, A. A., Holland, S., Kuin, P., & Pritchard, T. 2014, *ApSS*, **354**, 89
- Campbell, H., D'Andrea, C. B., Nichol, R. C., et al. 2013, *ApJ*, **763**, 88
- Childress, M., Aldering, G., Antilogus, P., et al. 2013, *ApJ*, **770**, 108
- Colless, M., Peterson, B. A., Jackson, C., et al. 2003, *arXiv:astro-ph/0306581*
- Conley, A., Guy, J., Sullivan, M., et al. 2011, *ApJS*, **192**, 1
- Contreras, C., Hamuy, M., Phillips, M. M., et al. 2010, *AJ*, **139**, 519
- Ergon, M., Jerkstrand, A., Sollerman, J., et al. 2015, *A&A*, **580**, A142
- Ergon, M., Sollerman, J., Fraser, M., et al. 2014, *A&A*, **562**, A17
- Fabricant, D., Fata, R., Roll, J., et al. 2005, *PASP*, **117**, 1411
- Falck, B. L., Riess, A. G., & Hlozek, R. 2010, *ApJ*, **723**, 398
- Foley, R. J., Challis, P. J., Chornock, R., et al. 2013, *ApJ*, **767**, 57
- Foley, R. J., & Mandel, K. 2013, *ApJ*, **778**, 167
- Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. 2013, *PASP*, **125**, 306
- Frieman, J. A., Bassett, B., Becker, A., et al. 2008, *AJ*, **135**, 338
- Ganeshalingam, M., Li, W., Filippenko, A. V., et al. 2010, *ApJS*, **190**, 418
- Garnavich, P. M., Bonanos, A. Z., Krisciunas, K., et al. 2004, *ApJ*, **613**, 1120
- Guillochon, J., Parent, J., Kelley, L. Z., & Margutti, R. 2017, *ApJ*, **835**, 64
- Gupta, R. R., Kuhlmann, S., Kovacs, E., et al. 2016, *AJ*, **152**, 154
- Guy, J., Astier, P., Baumont, S., et al. 2007, *A&A*, **466**, 11
- Guy, J., Sullivan, M., Conley, A., et al. 2010, *A&A*, **523**, A7
- Hamuy, M., Folatelli, G., Morrell, N. I., et al. 2006, *PASP*, **118**, 2
- Hlozek, R., Kunz, M., Bassett, B., et al. 2012, *ApJ*, **752**, 79
- Hsiao, E. Y., Conley, A., Howell, D. A., et al. 2007, *ApJ*, **663**, 1187
- Jerkstrand, A., Ergon, M., Smart, S. J., et al. 2015, *A&A*, **573**, A19
- Jones, D. H., Read, M. A., Saunders, W., et al. 2009, *MNRAS*, **399**, 683
- Jones, D. O. 2017, Measuring Dark Energy With Photometrically Classified Pan-STARRS Supernova: Bayesian Estimation Applied to Multiple Species Algorithm, <https://doi.org/10.5281/zenodo.572990>
- Kaiser, N., Burgett, W., Chambers, K., et al. 2010, *Proc. SPIE*, **7733**, 77330E
- Kelly, B. C. 2007, *ApJ*, **665**, 1489
- Kelly, P. L., Hicken, M., Burke, D. L., Mandel, K. S., & Kirshner, R. P. 2010, *ApJ*, **715**, 743
- Kessler, R., Bassett, B., Belov, P., et al. 2010, *PASP*, **122**, 1415
- Kessler, R., Becker, A. C., Cinabro, D., et al. 2009a, *ApJS*, **185**, 32
- Kessler, R., Bernstein, J. P., Cinabro, D., et al. 2009b, *PASP*, **121**, 1028
- Kessler, R., Marriner, J., Childress, M., et al. 2015, *AJ*, **150**, 172
- Kessler, R., & Scolnic, D. 2017, *ApJ*, **836**, 56
- Kunz, M., Bassett, B. A., & Hlozek, R. A. 2007, *PhRvD*, **75**, 103508
- Kurtz, M. J., & Mink, D. J. 1998, *PASP*, **110**, 934
- Lampeitl, H., Smith, M., Nichol, R. C., et al. 2010, *ApJ*, **722**, 566
- Le Fèvre, O., Vettolani, G., Garilli, B., et al. 2005, *A&A*, **439**, 845
- Leaman, J., Li, W., Chornock, R., & Filippenko, A. V. 2011, *MNRAS*, **412**, 1419
- Lewis, A., & Bridle, S. 2002, *PhRvD*, **66**, 103511
- Li, W., Leaman, J., Chornock, R., et al. 2011, *MNRAS*, **412**, 1441
- Lilly, S. J., Le Fèvre, O., Renzini, A., et al. 2007, *ApJS*, **172**, 70

- Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, *ApJS*, **225**, 31
- Lunnan, R., Chornock, R., Berger, E., et al. 2015, *ApJ*, **804**, 90
- March, M. C., Trotta, R., Berkes, P., Starkman, G. D., & Vaudrevange, P. M. 2011, *MNRAS*, **418**, 2308
- Marriner, J., Bernstein, J. P., Kessler, R., et al. 2011, *ApJ*, **740**, 72
- Metlova, N. V., Tsvetkov, D. Y., Shugarov, S. Y., Esipov, V. F., & Pavlyuk, N. N. 1995, *AstL*, **21**, 598
- Mink, D. J., Wyatt, W. F., Caldwell, N., et al. 2007, in ASP Conf. Ser. 376, *Astronomical Data Analysis Software and Systems XVI*, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco, CA: ASP), 249
- Modjaz, M., Blondin, S., Kirshner, R. P., et al. 2014, *AJ*, **147**, 99
- Modjaz, M., Li, W., Filippenko, A. V., et al. 2001, *PASP*, **113**, 308
- Möller, A., Ruhlmann-Kleider, V., Leloup, C., et al. 2016, *JCAP*, **12**, 008
- Newman, J. A., Cooper, M. C., Davis, M., et al. 2013, *ApJS*, **208**, 5
- Nugent, P., Kim, A., & Perlmutter, S. 2002, *PASP*, **114**, 803
- Oguri, M., & Marshall, P. J. 2010, *MNRAS*, **405**, 2579
- Pastorello, A., Kasliwal, M. M., Crockett, R. M., et al. 2008, *MNRAS*, **389**, 955
- Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, *ApJ*, **517**, 565
- Press, W. H. 1997, in *Unsolved Problems in Astrophysics*, ed. J. N. Bahcall & J. P. Ostriker (Princeton, NJ: Princeton Univ. Press), 49
- Rest, A., Scolnic, D., Foley, R. J., et al. 2014, *ApJ*, **795**, 44
- Rest, A., Stubbs, C., Becker, A. C., et al. 2005, *ApJ*, **634**, 1103
- Richardson, D., Jenkins, R. L., III, Wright, J., & Maddox, L. 2014, *AJ*, **147**, 118
- Richardson, D., Thomas, R. C., Casebeer, D., et al. 2001, *BAAS*, **33**, 1428
- Richmond, M. W., Treffers, R. R., Filippenko, A. V., & Paik, Y. 1996, *AJ*, **112**, 732
- Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, *AJ*, **116**, 1009
- Rodney, S. A., Riess, A. G., Strolger, L.-G., et al. 2014, *AJ*, **148**, 13
- Rubin, D., Aldering, G., Barbary, K., et al. 2015, *ApJ*, **813**, 137
- Sako, M., Bassett, B., Becker, A. C., et al. 2014, arXiv:1401.3317
- Sako, M., Bassett, B., Connolly, B., et al. 2011, *ApJ*, **738**, 162
- Sanders, N. E., Soderberg, A. M., Gezari, S., et al. 2015, *ApJ*, **799**, 208
- Scodeggio, M., Guzzo, L., Garilli, B., et al. 2016, arXiv:1611.07048
- Scolnic, D., & Kessler, R. 2016, *ApJL*, **822**, L35
- Scolnic, D., Rest, A., Riess, A., et al. 2014a, *ApJ*, **795**, 45
- Scolnic, D. M., Riess, A. G., Foley, R. J., et al. 2014b, *ApJ*, **780**, 37
- Shivvers, I., Mazzali, P., Silverman, J. M., et al. 2013, *MNRAS*, **436**, 3614
- Shivvers, I., Modjaz, M., Zheng, W., et al. 2017, *PASP*, **129**, 054201
- Silverman, J. M., Foley, R. J., Filippenko, A. V., et al. 2012, *MNRAS*, **425**, 1789
- Smee, S. A., Gunn, J. E., Uomoto, A., et al. 2013, *AJ*, **146**, 32
- Smith, M., Bacon, D. J., Nichol, R. C., et al. 2014, *ApJ*, **780**, 24
- Stritzinger, M. D., Phillips, M. M., Boldt, L. N., et al. 2011, *AJ*, **142**, 156
- Suh, H., Yoon, S.-c., Jeong, H., & Yi, S. K. 2011, *ApJ*, **730**, 110
- Sullivan, M., Guy, J., Conley, A., et al. 2011, *ApJ*, **737**, 102
- Sullivan, M., Le Borgne, D., Pritchet, C. J., et al. 2006, *ApJ*, **648**, 868
- Taubenberger, S., Hachinger, S., Pignata, G., et al. 2008, *MNRAS*, **385**, 75
- Taubenberger, S., Navasardyan, H., Maurer, J. I., et al. 2011, *MNRAS*, **413**, 2140
- Tonry, J., & Davis, M. 1979, *AJ*, **84**, 1511
- Tripp, R. 1998, *A&A*, **331**, 815
- Tsvetkov, D. Y., Volkov, I. M., Baklanov, P., Blinnikov, S., & Tuchin, O. 2009, *PZ*, 29
- Yaron, O., & Gal-Yam, A. 2012, *PASP*, **124**, 668